КОНСПЕКТ ЛЕКТОРА (продолжение)
математический анализ, 1 курс, 4 модуль, 2016, А.М. Красносельский
Лекция 1 (04 апреля 2016)

1 Собственные интегралы, зависящие от параметра

Теперь мы начнём новую тему, интегралы, зависящие от параметра. Такими интегралами описываются некоторые (многие) важные вещи, в частности, гамма-функция, та, которая факториал.

Эта тема содержит 2 части. Первая — это абстрактные теоремы. Есть функция, которая определяется в виде интеграла, зависящего от параметра. Надо знать про неё, когда она непрерывная, когда она дифференцируемая, чему равна производная.

А вторая часть — это что такое самая важная такая функция, гамма-функция.

Пусть $y \in Y = [c,d]$, $a(y), b(y) \in C[c,d]$, пусть при каждом y функция $f(x,y)$ (как функция переменной x) интегрируема (непрерывна) на промежутке $[a(y), b(y)]$. Тогда определена функция

$$\Phi(y) = \int_{a(y)}^{b(y)} f(x,y)dx, \quad \text{часто } a(y) \text{ и } b(y) \text{ — постоянные:} \quad \Phi(y) = \int_a^b f(x,y)dx.$$

— это и есть собственный интеграл, зависящий от параметра.

Важное соображение такое. Вот у вас есть интеграл $I(a,b,y) = \displaystyle\int_a^b f(t,y)dt$, здесь ровно 3 места, куда можно вставить параметры. Можно считать, что это у нас такая функция от трёх переменных. Соответственно, $\Phi(y) = I(a(y), b(y), y)$ — это композиция функции I и функций a и b.

Непрерывность, дифференцируемость. Мы считаем, что $a(y) < b(y)$, однако, это не важно, просто, чтобы картинку рисовать естественнее.

Обозначим $G = \{(x,y) : y \in [c,d], x \in [a(y), b(y)]\}$.

Теорема. $a(y), b(y) \in C[c,d]$, $f(x,y) \in C(G) \Rightarrow \Phi(y) \in C[a,b]$.

Доказательство. Пишем

$$\begin{aligned}
|\Phi(y+h) - \Phi(y)| &= \Big| \int_{a(y+h)}^{b(y+h)} f(x,y+h)dx - \int_{a(y)}^{b(y)} f(x,y)dx \Big| \\
&\leqslant \Big| \int_{a(y+h)}^{b(y+h)} f(x,y+h)dx - \int_{a(y)}^{b(y)} f(x,y+h)dx \Big| + \Big| \int_{a(y)}^{b(y)} f(x,y+h)dx - \int_{a(y)}^{b(y)} f(x,y)dx \Big| \\
&\leqslant (|a(y+h) - a(y)| + |b(y+h) - b(y)|) \sup_{x,y} |f| + |b(y) - a(y)| \sup_{x,y,h} |f(x,y+h) - f(x,y)|.
\end{aligned}$$

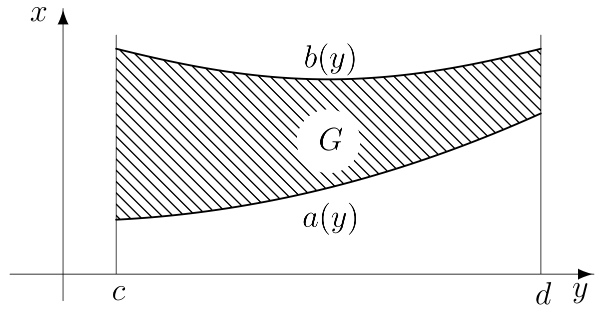


Рис. 1: Область G на плоскости $\{x, y\}$

Малость первых двух слагаемых следует из непрерывности функций a и b, малость третьего слагаемого следует из равномерной непрерывности функции двух переменных f по теореме Кантора (G — компакт). $\square$

Следствие (перестановка интеграла и предела). *Пусть* $c \in [a, b] \Rightarrow G$ *— прямоугольник,*

$$\lim_{y\to c}\int_a^b f(x,y)dx = \int_a^b \lim_{y\to c} f(x,y)dx = \int_a^b f(x,c)dx.$$

Здесь пределы интегрирования должны быть постоянные, чтобы y не вылез за знак предела.

Теорема. $a(y), b(y) \in C^1[c,d],\ f(x,y) \in C(G)\ \exists f'_y(x,y) \in C(G)\ \Rightarrow\ \Phi(y) \in C^1[a,b]$ и

$$\Phi'(y) = f(b(y),y)b'(y) - f(a(y),y)a'(y) + \int_{a(y)}^{b(y)} f'_y(x,y)dx.$$

Доказательство. В лоб: пишем $\Delta\Phi(y)/\Delta y$ и расписываем по кусочкам. К кусочку, порождающему слагаемое с интегралом, применяем теорему о непрерывности по Δy, к остальным двум — теорему о среднем, всё получится. Трудный кусок:

$$A = \left|\int_{a(y)}^{b(y)} \left(f(x,y+h) - f(x,y) - h\,\frac{\partial f}{\partial y}(x,y)\right)\, dx\right| \leqslant \int_{a(y)}^{b(y)} \left|f(x,y+h) - f(x,y) - h\,\frac{\partial f}{\partial y}(x,y)\right|\, dx.$$

Подынтегральное выражение J в последнем интеграле удовлетворяет неравенству

$$J \leqslant |h| \sup_{\theta\in(0,1)} \left|\frac{\partial f}{\partial y}(x, y+\theta h) - \frac{\partial f}{\partial y}(x,y)\right|.$$

Последнее выражение мало при малых h по теореме Кантора, поэтому $A = o(h)$. $\square$

Перестановка интегралов. Пусть задана функция $f(x,y) \in C([a,b]\times[c,d])$. Тогда функция $\Phi(x) = \displaystyle\int_c^d f(x,y)\,dy$ определена и непрерывна на $[a,b]$ и её можно интегрировать. Интеграл от этой функции называется *повторным* интегралом, обозначается

$$\int_a^b \left(\int_c^d f(x,y)dy \right) dx \qquad \text{или} \qquad \int_a^b dx \int_c^d f(x,y)\,dy.$$

Аналогично можно определить другой повторный интеграл

$$\int_c^d \left(\int_a^b f(x,y)dx \right) dy = \int_c^d dy \int_a^b f(x,y)\,dx.$$

Теорема. *Пусть* $f(x,y) \in C([a,b] \times [c,d])$. *Тогда* $\displaystyle\int_a^b dx \int_c^d f(x,y)\,dy = \int_c^d dy \int_a^b f(x,y)\,dx.$

Доказательство. 1) Рассмотрим функцию $F(t,y) = \displaystyle\int_a^t f(x,y)dx$.

2) $F \in C([a,b] \times [c,d])$, это следует из оценки

$$\Big| \int_a^{t+\Delta} f(x,y+h)dx - \int_a^t f(x,y)dx \Big| \leqslant |\Delta| \sup |f| + |b-a| \sup |f(x,y+h) - f(x,y)|.$$

3) Положим, $F_1(t) = \displaystyle\int_a^t \left(\int_c^d f(x,y)dy \right) dx, \quad F_2(t) = \int_c^d \left(\int_a^t f(x,y)dx \right) dy.$

Продифференцируем их:

$$F_1'(t) = \int_c^d f(t,y)dy, \qquad F_2'(t) = \int_c^d f(t,y)dy \;\Rightarrow\; F_1(t) - F_2(t) \equiv const.$$

4) Теперь $F_1(a) = F_2(a) = 0 \;\Rightarrow\; F_1(b) = F_2(b)$. □

Если бы мы знали, что такое двойной интеграл, то мы бы сказали ещё, что

$$\iint_{[a,b]\times[c,d]} f(x,y)\,dxdy = \int_a^b dx \int_c^d f(x,y)\,dy = \int_c^d dy \int_a^b f(x,y)\,dx.$$

Функция f предполагается непрерывной, чтобы функция Φ была интегрируемой. Общие условия, при которых существует повторный интеграл Римана хитрые, есть в книжках, я их не буду доказывать. Примерно они выглядят так.

Пусть функция f двух переменных интегрируема в прямоугольнике и пусть для каждого y существует однократный интеграл по x. Тогда $\exists$ соответствующий повторный интеграл.

Доказательство не сложное, но используются двойные интегралы.

2 Несобственные интегралы, зависящие от параметра

Пусть $y \in Y$, пусть при каждом y функция $f(x,y)$ (как функция переменной x) интегрируема (непрерывна) на промежутке $[a,\infty)$. Тогда определена функция $\Phi(y) = \displaystyle\int_a^\infty f(x,y)dx$, — это и есть несобственный интеграл, зависящий от параметра.

Заметим, что у нас сходимости несобственного интеграла при разных y никак не связаны. То есть, это типичная поточечная сходимость: при каждом y сходится интеграл.

Бывают ещё аналогичные интегралы $\Phi(y) = \int_a^b f(x,y)dx$, в которых при некоторых или при всех значениях y функция f неограниченная. Специально мы такие интегралы изучать не будем, заменой переменных их можно свести к случаю неограниченного промежутка, идейных особенностей нет.

Очень многие факты о функции $\Phi(y)$ совпадают с фактами о функциях $\sum\limits_{n=1}^{\infty} a_n(y),\ \lim\limits_{n\to\infty} s_n(y)$.

Равномерная сходимость. Определение равномерной сходимости к $\Phi(y)$:

$$\forall\ \varepsilon > 0\ \exists\ w = w(\varepsilon): \quad \forall\ b > w,\ y \in Y \quad \left|\int_a^b f(x,y)dx - \Phi(y)\right| < \varepsilon.$$

Поточечная сходимость:

$$\forall y \in Y\ \forall\ \varepsilon > 0\ \exists\ w = w(\varepsilon, y): \quad \forall\ b > w, \quad \left|\int_a^b f(x,y)dx - \Phi(y)\right| < \varepsilon.$$

Теорема (Критерий Коши равномерной сходимости).

$$\forall\ \varepsilon > 0\ \exists\ w = w(\varepsilon): \quad \forall\ b_1, b_2 > w,\ \forall y \in Y \quad \left|: \int_{b_1}^{b_2} f(x,y)dx\right| < \varepsilon.$$

Доказательство. 1) В одну сторону совсем просто:

$$\left|\int_{b_1}^{b_2} f(x,y)dx\right| \le \left|\int_{b_1}^{\infty} f(x,y)dx - \int_{b_2}^{\infty} f(x,y)dx\right| \le \left|\int_{b_1}^{\infty} f(x,y)dx - \Phi(y)\right| + \left|\int_{b_2}^{\infty} f(x,y)dx - \Phi(y)\right|$$

2) В другую сторону: Φ определена из-за обычного критерия Коши, не равномерного

$$\int_a^{\infty} f(x,y)dx = \Phi(y).$$

Теперь из Коши следует $\left|\int_{b_1}^{\infty} f(x,y)dx\right| \leqslant \varepsilon\ (\forall y, b_1 > w) \quad \Rightarrow \quad \left|\int_a^{b_1} f(x,y)dx - \Phi(y)\right| \le \varepsilon.$ □

Пример использования признака Коши. $y \in [0, \infty)$

$$G(y) = \int_0^{\infty} \frac{dx}{1 + (x-y)^2}$$

Сходится при всех y, но неравномерно. Сходимость очевидна. Неравномерность следует из признака Коши:

$$\int_{b_1}^{b_2} \frac{dx}{1 + (x-y)^2} = \operatorname{arctg}(b_1 - y) - \operatorname{arctg}(b_2 - y)$$

Зафиксируем $\varepsilon > 0$ (например, .1), по сколь угодно большим b_1, b_2 найдется y такой (например, $b_2 = y = b_1 + 1$, что это по модулю равно $\operatorname{arctg} 1 > \varepsilon$.

Лекция 2 (11 апреля 2016)

На прошлой лекции мы рассмотрели собственные интегралы, зависящие от параметра, их непрерывность, как их дифференцировать, как переставлять два повторных интеграла.

Начали рассматривать несобственные интегралы, зависящие от параметра, их равномерную сходимость, провели аналогию с функциональными рядами и последовательностями функций, сформулировали и доказали критерий Коши равномерной сходимости.

Равномерная абсолютная сходимость — если интеграл от модуля равномерно сходится, то интеграл от функции также равномерно сходится.

Теорема (мажорантный признак). *Пусть даны две функции $f_1(x,y) \geqslant f_2(x,y) \geqslant 0$, обе непрерывные на $[a,\infty)$. Тогда из равномерной сходимости несобственного интеграла от большей следует равномерная сходимость интеграла от меньшей.*

Признак следует из критерия Коши. Сначала напишем критерий Коши равномерной сходимости для $\displaystyle\int^{\infty} f_1(x,y)dx$, из него следует критерий Коши для равномерной сходимости для $\displaystyle\int^{\infty} f_2(x,y)dx$. Отсюда следует равномерная и абсолютная сходимость для этого интеграла. $\square$

Непрерывность можно не предполагать, достаточно, чтобы функции были интегрируемы на каждом конечном промежутке. Иными словами, чтобы множество точек разрыва функций было множеством меры ноль по Лебегу.

Отсюда следует главный признак абсолютной равномерной сходимости. Главный — в смысле часто используемый.

Теорема (Признак Вейерштрасса, мажорантный). *Если $\forall y$ 1) функция $f(x,y)$ интегрируема на любом отрезке $[a,b] \in [a,\infty)$ и 2) $|f(x,y)| \leqslant g(x)$, причём $\displaystyle\int^{\infty} g(x)dx < \infty$, то несобственный интеграл $\displaystyle\int^{\infty} f(x,y)dx$ сходится абсолютно и равномерно.*

Пример применения признака Вейерштрасса. Пусть $y \in [\delta, 1],\ \delta > 0$,

$$\int_1^{\infty} \frac{dx}{x^{y+1}},$$

Очевидно, интеграл сходится равномерно: $x^{y+1} \geqslant x^{1+\delta}$, интеграл от $x^{-1-\delta}$ сходится.

Признаки Абеля и Дирихле равномерной сходимости.

Пусть функции f, g определены и непрерывны на $[a,\infty) \times Y$, функция f равномерно ограничена, $\forall y \in Y$ функция $f(x,y)$ монотонная, как функция переменной x.

Дирихле: $\displaystyle\int_a^u g(x,y)dx$ равномерно по u, y ограничено, $\displaystyle\lim_{x\to\infty} f(x,y) = 0$ равномерно по $y \Rightarrow$ равномерно сходится $\displaystyle\int_a^{\infty} f(x,y)g(x,y)dx$.

Абель: $\int_a^\infty g(x,y)dx$ равномерно сходится $\Rightarrow$ равномерно сходится $\int_a^\infty f(x,y)g(x,y)dx$.

Доказательства обоих признаков следуют из второй теоремы о среднем и критерия Коши. Если функция f монотонная при каждом $y \in Y$, как функция от переменной x, то $\exists \xi \in [a,b]$

$$\int_a^b f(x,y)g(x,y)dx = f(a,y)\int_a^\xi g(x,y)dx + f(b,y)\int_\xi^b g(x,y)dx, \qquad \xi = \xi(y).$$

Теперь перейдем к интегрированию и дифференцированию несобственных интегралов, зависящих от параметров. Ситуация и результаты очень близки к тому, что было с рядами.

Здесь есть 2 методических сценария. Можно честно передоказывать все теоремы, а я «пойду другим путём». Буду считать, что мы всё знаем про ряды и последовательности, сведу все необходимые теоремы к последовательностям. Точно так же, как пределы функций можно свести к пределам последовательностей, воспользовавшись пределами по Гейне или секвенциальной сходимостью, это одно и то же.

Секвенциальная равномерная сходимость (по Гейне) интегралов.

Аналогий с последовательностями и рядами очень много. Чтобы пользоваться уже доказанными теоремами, можно применить пределы по Гейне, как когда-то в первом семестре.

Определение. *Будем говорить, что несобственный интеграл*

$$\Phi(y) = \int_a^\infty f(x,y)dx$$

равномерно сходится по Гейне, если $\forall$ монотонной последовательности $A_n \to \infty$ последовательность функций $\int_a^{A_n} f(x,y)dx$ равномерно сходится к $\Phi(y)$.

Очевидно, что предел общий для всех последовательностей.

Замечание. Несобственный интеграл равномерно сходится по Гейне, если $\forall$ монотонной последовательности $A_n \to \infty$ ряд $\sum_{n=1}^\infty \int_{A_n}^{A_{n+1}} f(x,y)dx$ равномерно сходится к $\Phi(y)$.

Теорема. *Интеграл $\Phi(y)$ равномерно сходится iff он равномерно сходится по Гейне.*

Доказательство. 1) Из равномерной сходимости следует равномерная сходимость по Гейне. Очевидно. Почему это очевидно? Напишем определение равномерной сходимости интеграла

$$\forall\ \varepsilon > 0\ \exists\ W = W(\varepsilon):\ \ \forall\ A > W,\ y \in Y\quad \left|\int_a^A f(x,y)dx - \Phi(y)\right| < \varepsilon.$$

Теперь выберем $A_n \to \infty$ и напишем определение равномерной сходимости последовательности функций $\int_a^{A_n} f(x,y)dx$ к $\Phi(y)$:

$$\forall\ \varepsilon > 0\ \exists\ N = N(\varepsilon):\ \ \forall\ n > N,\ y \in Y\quad \left|\int_a^{A_n} f(x,y)dx - \Phi(y)\right| < \varepsilon.$$

Теперь совсем очевидно.

2) Из равномерной сходимости по Гейне следует равномерная сходимость. Рассуждаем от противного. Пусть есть равномерная сходимость по Гейне, но нет равномерной сходимости. Раз нет равномерной сходимости, то не выполняется признак Коши равномерной сходимости интеграла:

$$\exists \varepsilon_0 > 0 \; \forall \; A > 0 \; \exists y, B_1 > B_2 > A : \left| \int_{B_2}^{B_1} f(x,y)dx \right| \geqslant \varepsilon_0.$$

Построим теперь такую последовательность $A_n \to \infty$ для которой нет равномерной сходимости функций $\displaystyle\int_a^{A_n} f(x,y)dx$, даже такой, что $\displaystyle\int_{A_n}^{A_{n+1}} f(x,y)dx \not\rightrightarrows 0$. Выберем A_0, по $A = A_0$ построим B_1, B_2 и y, положим $A_1 = B_2, A_2 = B_1, y_1 = y$. Потом по $A = B_1$ снова построим $B_1, B_2 > A_2$ и y, положим $A_3 = B_2, A_4 = B_1, y_2 = y$. Получится последовательность y_n, A_n такую, что $\displaystyle\left| \int_{A_{2n+1}}^{A_{2n}} f(x,y_n)dx \right| \geqslant \varepsilon_0$. Это противоречит равномерной фундаментальности последовательности функций $\displaystyle\int_a^{A_n} f(x,y)dx$. Противоречие. □

Про монотонность последовательности A_n. Можно было бы слово монотонный в определении сходимости по Гейне опустить... но удобнее оставить!

Когда-то была доказана такая теорема: пусть c — предельная точка множества G, пусть

$$F_n(y) \overset{G}{\rightrightarrows} F(y), \; \exists \lim_{y \to c} F_n(y) = a_n \quad \Rightarrow \quad a_n \to b, \; \lim_{y \to c} F(y) = b.$$

Надо так считать, что G — это полуоткрытый промежуток, с концом $c \neq G$.

Из неё следует естественный аналог для интегралов: $A_n \to \infty$, монотонная,

$$\int_a^{A_n} f(x,y)\,dx \rightrightarrows F(y), \; \exists \lim_{y \to c} \int_a^{A_n} f(x,y)\,dx = a_n \quad \Rightarrow \quad a_n \to b, \; \lim_{y \to c} F(y) = b.$$

Первое условие — это равномерная сходимость на G интеграла. Второе условие, если функции $f(x,y)$ непрерывны при $x \in [a,\infty), y \in G \bigcup c$, то

$$\lim_{y \to c} \int_a^{A_n} f(x,y)\,dx = \int_a^{A_n} f(x,c)\,dx = a_n.$$

Это была такая теорема (следствие), что предел можно переставлять с собственным интегралом, если внутри всё непрерывно.

То есть, если интеграл на G равномерно сходился и функции были непрерывны на $x \in [a,\infty), y \in G \bigcup c$, то существует $\lim a_n = \displaystyle\int_a^{\infty} f(x,c)\,dx$ и

$$\int_a^{\infty} f(x,c)\,dx = \lim_{y \to c} \int_a^{\infty} f(x,y)\,dx.$$

Иными словами, если $f(x,y)$ непрерывны на $x \in [a,\infty), y \in [b,c]$, несобственный интеграл равномерно сходится к $F(y)$ на $G = [b,c)$, то можно менять предел и интеграл местами, сходится интеграл при $y = c$.

Теперь сразу получаем равномерную непрерывность предела на замкнутом промежутке $y \in [b,c]$, непрерывность предела и можно переставлять предел и несобственный интеграл.

Наоборот, если при $y = \alpha$ интеграл расходится, или получился разрывный предел (то есть нельзя переставлять), то на незамкнутом промежутке не может быть равномерной сходимости.

Пример. Рассмотрим пример $\displaystyle\int_1^\infty \frac{y\,dx}{x^{1+y}}$. Покажем, что интеграл не сходится равномерно на промежутке $y \in [0,1]$, но есть неравномерная поточечная сходимость. Предел — разрывная функция.

Итак, $y \in [0,1]$, $\displaystyle\int_1^\infty yx^{-1-y}\,dx$. Очевидно, интеграл сходится поточечно, для сходимости при $y = 0$ добавлен множитель y. Этот интеграл равен 0 при $y = 0$ и равен 1 при $y > 0$:

$$\int_1^\infty yx^{-1-y}\,dx = x^{-y}\Big|_1^\infty = 1.$$

Равномерной сходимости нет — предел разрывен. Поэтому нельзя переставлять предел с интегралом:

$$\lim_{y\to 0}\int_1^\infty yx^{-1-y}\,dx = 1 \neq 0 = \int_1^\infty \lim_{y\to 0}(yx^{-1-y})\,dx.$$

Вот теорема про то, что **равномерно сходящийся интеграл сходится к непрерывной функции**, точно так же как равномерный предел последовательности непрерывных функций сходится к непрерывной функции.

Напомнить про равномерную сходимость функциональных последовательностей и теорему: *Если последовательность непрерывных функций равномерно сходится, то предел непрерывен.* Можно воспользоваться равномерной сходимостью по Гейне и свести к пределу последовательности, я отдельно докажу в лоб.

Теорема. $f(x,y) \in C([a,\infty)\times[\alpha,\beta])$; $\displaystyle\int_a^\infty f(x,y)dx \rightrightarrows F(y) \quad \Rightarrow \quad F(y) \in C([\alpha,\beta])$

Доказательство. Сначала разбить $|F(y+\Delta y) - F(y)|$ на 3 части, написать оценку

$$\Big|F(y+\Delta y) - F(y)\Big| \leqslant \left|F(y+\Delta y) - \int_a^A f(x,y+\Delta y)dx\right| + \left|F(y) - \int_a^A f(x,y)dx\right| + \\ + \left|\int_a^A f(x,y+\Delta y)dx - \int_a^A f(x,y)dx\right|.$$

Потом выберем число A такое, чтобы 2 первых слагаемых были маленьких, и зафиксируем его. Далее, при этом фиксированном A воспользуемся равномерной непрерывностью функции f на компакте — прямоугольнике $[a,A]\times[\alpha,\beta]$ (теоремой Кантора). $\square$

Заметим, что последовательность именно такая: равномерной непрерывности на полуполосе может и не быть. Заметим также, что вместо $Y = [\alpha, \beta]$ может быть выбран любой компакт.

В прошлом модуле была теорема Дини (я её сформулирую для рядов): если ряд из неотрицательных непрерывных функций на отрезке поточечно сходится к непрерывной функции, то он сходится равномерно. Справедлив полный аналог этой теоремы и для интегралов.

Теорема Дини для несобственных интегралов. *Пусть $f(x,y) \geqslant 0$ при $x \in [a, \infty), y \in [\alpha, \beta]$, пусть функции $f(x,y)$ непрерывны и пусть при каждом y интеграл*

$$\Phi(y) = \int_a^\infty f(x,y)dx$$

сходится, причём предельная функция Φ непрерывна на $y \in [\alpha, \beta]$. Тогда сходимость интеграла равномерная.

Теорема прямо вытекает из сходимости по Гейне. Берем любую монотонную последовательность A_n, из неё и теоремы Дини для рядов получаем равномерную сходимость монотонной последовательности непрерывных функций $\displaystyle\int_a^{A_n} f(x,y)dx$, отсюда следует равномерная сходимость интеграла. $\square$

3 Интегрирование и дифференцирование несобственных интегралов, зависящих от параметра

Предыдущие теоремы — о перестановке предела и интеграла. Теперь займемся перестановками интеграла и дифференцирования и интегрирования.

Теорема (интегрирование). *Пусть $f(x,y) \in C([a,\infty) \times [\alpha, \beta])$,*

$$\int_a^\infty f(x,y)dx \rightrightarrows F(y) \quad \text{на } [\alpha, \beta] \quad \Rightarrow \qquad \int_\alpha^\beta F(y)dy = \int_a^\infty \left(\int_\alpha^\beta f(x,y)dy \right) dx$$

Доказательство. Выберем и зафиксируем $A_n \to \infty$. Положим

$$F_n(y) = \int_a^{A_n} f(x,y)dx \rightrightarrows F(y) \quad \Rightarrow \quad \int_\alpha^\beta F_n(y)dy \to \int_\alpha^\beta F(y)dy.$$

Но $\displaystyle\int_\alpha^\beta F_n(y)dy = \int_\alpha^\beta \left(\int_a^{A_n} f(x,y)dx \right) dy = \int_a^{A_n} \left(\int_\alpha^\beta f(x,y)dy \right) dx$

$$\Rightarrow \quad \lim_{n\to\infty} \int_a^{A_n} \left(\int_\alpha^\beta f(x,y)dy \right) dx = \int_\alpha^\beta F(y)dy \quad \Rightarrow \quad \square$$

Лекция 3 (15 апреля 2016)

На прошлой лекции мы перешли к функциям, заданным в виде несобственных интегралов с параметром:

1) Сформулировали признаки абсолютной равномерной сходимости: мажорантный и Вейерштрасса, почти как для рядов.

2) Сформулировали признаки равномерной сходимости Дирихле и Абеля, тоже как для рядов.

3) Сформулировали утверждение про равномерную сходимость по Гейне, очень похоже как для чисел.

4) Сформулировали теорему о перестановке предела и несобственного интеграла.

5) Сформулировали теорему о непрерывности предела равномерно сходящегося интеграла от непрерывных функций.

6) Сформулировали теорему Дини для интегралов.

7) Сформулировали теорему о перестановке собственного и несобственного интегралов.

Последняя теорема звучала примерно так: если несобственный интеграл от непрерывной функции $f(x, y)$ равномерно сходится, его можно переставлять с обычным собственным интегралом.

Пример. Пусть $f(x, y) = (2 - xy)xye^{-xy}$ на $x \in [0, +\infty)$, $y \in [0, 1]$. Очевидно, что функция f на этом множестве непрерывна. Попробуем посчитать повторные интегралы в разном порядке, при этом мы используем равенство $(u^2e^{-u})' = (2 - u)ue^{-u}$. С одной стороны,

$$\int_0^\infty dx \int_0^1 (2 - xy)xye^{-xy}\, dy = \int_0^\infty \frac{dx}{x} \int_0^x (2 - u)ue^{-u}\, du = \int_0^\infty \frac{dx}{x}\ u^2e^{-u}\Big|_0^x = \int_0^\infty xe^{-x}\, dx = 1 \neq 0.$$

С другой:

$$\int_0^1 dy \int_0^\infty (2 - xy)xye^{-xy}\, dx = \int_0^1 \frac{dy}{y} \int_0^\infty (2 - u)ue^{-u}\, du = \int_0^1 \frac{dy}{y}\ u^2e^{-u}\Big|_0^\infty = 0.$$

Это пример о том, что равномерная сходимость существенная. Здесь есть поточечная сходимость — мы же посчитали все интегралы, всё получилось. Громоздкость формул... некоторая... определялась тем, что хотелось, чтобы все в явном виде интегрировалось. Отдельно добавлю, что совсем простым пример и не мог быть. Иначе по теореме Дини получилась бы равномерная сходимость.

То, что нет равномерной на $(0, 1)$ сходимости $\displaystyle\int_0^\infty f(x, y)\, dx$, почти очевидно: величина xy принимает любые значения при больших x и малых y.

Из доказанной теоремы о перестановке интеграла и несобственного интеграла следует интересное следствие, использующее теорему Дини.

Пусть $f(x,y) \geqslant 0$ непрерывная и пусть несобственный интеграл сходится поточечно к непрерывной функции. Тогда по теореме Дини есть сходимость равномерная и можно переставлять интегралы местами.

Теперь перейдем к производным. У нас была теорема: если последовательность $F_n(y)$ сходилась поточечно (или хотя бы в одной точке), а последовательность производных $F_n'(y)$ сходилась равномерно, то можно было переставлять производную и предел:

$$\frac{d}{dy}\Big(\lim_{n\to\infty} F_n(y)\Big) = \lim_{n\to\infty} F_n'(y).$$

Из этой теоремы следует её аналог для интегралов.

Теорема (дифференцирование). *Пусть* $f(x,y),\ f_y'(x,y) \in C([a,\infty)\times[\alpha,\beta])$,

$$\int_a^\infty f(x,y)dx \to F(y), \quad \int_a^\infty f_y'(x,y)dx \overset{[\alpha,\beta]}{\rightrightarrows} G(y) \qquad \Rightarrow \qquad \exists F'(y) = G(y).$$

Доказательство. Выберем $A_n \to \infty$, тогда при $n \to \infty$

$$\int_a^{A_n} f_y'(x,y)dx \rightrightarrows G(y) \quad \text{и} \quad \int_a^{A_n} f(x,y)dx \to F(y).$$

По теореме о дифференцировании собственного интеграла: $\exists \left(\int_a^{A_n} f(x,y)dx\right)_y' = \int_a^{A_n} f_y'(x,y)dx.$

По теореме о дифференцировании пределов функциональных последовательностей:

$$G(y) = \lim_{n\to\infty}\int_a^{A_n} f_y'(x,y)dx = \lim_{n\to\infty}\left(\int_a^{A_n} f(x,y)dx\right)_y' = \left(\lim_{n\to\infty}\int_a^{A_n} f(x,y)dx\right)_y' = F_y'(y),$$

следовательно, $\exists F'(y) = G(y)$. □

Теорема 1 о перестановке несобственных интегралов. *Пусть функция $f(x,y)$ непрерывна на $[a,\infty)\times[b,\infty)$ и пусть интегралы*

$$\int_a^\infty f(x,y)\,dx = G(y), \qquad \int_b^\infty f(x,y)\,dy = H(x)$$

равномерно сходятся на каждом конечном промежутке $x \in [a,A]$ или $y \in [b,B]$ значений параметра. Пусть повторный интеграл

$$\int_a^\infty dx \int_b^\infty |f(x,y)|\,dy$$

сходится. Тогда определен второй повторный интеграл и оба повторных интеграла равны:

$$\exists \int_a^\infty H(x)\,dx = \int_b^\infty G(y)\,dy \qquad \Leftrightarrow \qquad \int_b^\infty dy \int_a^\infty f(x,y)\,dx = \int_a^\infty dx \int_b^\infty f(x,y)\,dy.$$

Доказательство. Сначала выберем большое значение d и положим $H_d(x) = \int_b^d f(x,y)dy$. Функция $H_d(x)$ определена при каждом d, непрерывна по $x \in [a, \infty)$. По предположению $\lim\limits_{d\to\infty} H_d(x) = H(x)$ при всех x равномерно на каждом конечном отрезке изменения x, отсюда следует непрерывность $H(x)$. Так как

$$|H_d(x)| = |\int_b^d f(x,y)\,dy| \leqslant \int_b^\infty |f(x,y)|\,dy = \Phi(x)$$

и по предположению интеграл $\int_a^\infty \Phi(x)\,dx$ сходится, то по мажорантному признаку интеграл $\int_a^\infty H_d(x)\,dx$ сходится равномерно по всем d.

Значит, в формуле $\lim\limits_{d\to\infty}\int_a^\infty H_d(x)\,dx$ можно менять предел и интеграл местами:

$$\lim_{d\to\infty}\int_a^\infty H_d(x)\,dx = \int_a^\infty \lim_{d\to\infty} H_d(x)\,dx = \int_a^\infty H(x)\,dx.$$

Теперь переставим местами собственный и несобственный интегралы в формуле

$$\int_a^\infty H_d(x)\,dx = \int_a^\infty dx\int_b^d f(x,y)dy = \int_b^d dy\int_a^\infty f(x,y)dx = \int_b^d G(y)\,dy$$

и сразу получим требуемое равенство. Переставлять собственный и несобственный интегралы при фиксированном d можно по доказанной теореме и предположению о равномерной сходимости на каждом конечном интервале. □

Теорема 2 о перестановке несобственных интегралов. *Пусть функция $f(x,y) \geqslant 0$ непрерывна на $[a,\infty)\times[b,\infty)$ и существуют интегралы*

$$\int_a^\infty f(x,y)dx = G(y) \in C([b,\infty)), \qquad \int_b^\infty f(x,y)dy = H(x) \in C([a,\infty)), \qquad \int_b^\infty G(y)dy.$$

Тогда определен второй повторный интеграл и оба повторных интеграла равны:

$$\exists \int_a^\infty H(x)dx = \int_b^\infty G(y)dy.$$

То есть снова можно переставлять интегралы:

$$\int_a^\infty dx\int_b^\infty f(x,y)dy = \int_b^\infty dy\int_a^\infty f(x,y)dx.$$

Теорема 2 следует из теоремы 1 в силу теоремы Дини. Из положительности подынтегральной функции и всех непрерывностей следует равномерная сходимость. Обращаю внимание, что в теореме Дини промежуток значений должен был быть компактен. Иначе все не верно (привести пример). Но в условии теоремы 1 требовалась равномерность именно на каждом конечном промежутке.

Пример 1. Пусть $f(x) = \cos 2x$ на отрезке $[-\pi/4, \pi/4]$, равна 0 в остальных точках отрезка $[-1, 1]$ и периодична по x с периодом 2. Пусть $g_n(x) = 0$ при $x < n$ и $g_n(x) = 1$ при $x \geqslant n$. Тогда последовательности непрерывных функций $f(x)g_n(x)$ при $x \geqslant 0$ монотонно поточечно сходится к непрерывной функции «тождественный ноль», но равномерной сходимости нет. Это пример о том, что в теореме Дини компактность множества значения параметра существенна.

Пример 2. Это пример, когда переставлять интегралы нельзя:

$$\int_1^\infty \frac{x^2 - y^2}{(x^2 + y^2)^2} dx = \frac{-x}{x^2 + y^2}\Big|_1^\infty = \frac{1}{1 + y^2}, \qquad \int_1^\infty \frac{x^2 - y^2}{(x^2 + y^2)^2} dy = \frac{-1}{1 + x^2},$$

поэтому

$$\int_1^\infty dy \int_1^\infty \frac{x^2 - y^2}{(x^2 + y^2)^2} dx = \frac{\pi}{4}, \quad \int_1^\infty dx \int_1^\infty \frac{x^2 - y^2}{(x^2 + y^2)^2} dy = -\frac{\pi}{4}.$$

В этом примере двойной интеграл $\displaystyle\iint_{1 \leqslant x, y < \infty} f(x, y)\, dxdy$ не существует, интеграл от модуля равен бесконечности.

Пример 3. Интеграл Эйлера–Пуассона:

$$J = \int_0^\infty e^{-u^2} du = y \int_0^\infty e^{-(xy)^2} dx, \ \ y > 0 \qquad (J = \frac{1}{2}\sqrt{\pi}).$$

Теперь положим $f(x, y) = ye^{-y^2(1+x^2)}$. С одной стороны

$$\int_0^\infty dy \int_0^\infty f(x, y)\, dx = \int_0^\infty dy\, e^{-y^2} \int_0^\infty ye^{-(xy)^2} dx = \int_0^\infty e^{-y^2} dy \ \int_0^\infty e^{-u^2} du = J^2.$$

С другой стороны,

$$\int_0^\infty dx \int_0^\infty f(x, y)\, dy = \int_0^\infty dx \int_0^\infty ye^{-y^2(1+x^2)} dy = \frac{1}{2} \int_0^\infty \frac{dx}{1 + x^2} = \frac{\pi}{4}.$$

Если бы обосновать перестановку интегралов, мы получили бы требуемое равенство $J = \frac{1}{2}\sqrt{\pi}$.

Как проверить перестановку? Есть принципиальная трудность:

$$\int_0^\infty f(x, y)\, dx = \int_0^\infty ye^{-y^2(1+x^2)}\, dx = \begin{cases} Je^{-y^2}, & y > 0; \\ 0, & y = 0. \end{cases}$$

Поэтому равномерной сходимости при $y \in [0, \delta]$ не может быть. Эту проблему можно обойти следующим образом. Пусть $\delta > 0$ малое число, рассмотрим интегралы

$$\int_\delta^\infty dy \int_0^\infty ye^{-y^2(1+x^2)}\, dx = \int_\delta^\infty e^{-y^2} dy \ \int_0^\infty e^{-u^2} du = J^2 + O(\delta)$$

и
$$\int_0^\infty dx \int_\delta^\infty ye^{-y^2(1+x^2)}dy = \int_0^\infty \frac{1}{2+2x^2}e^{-\delta^2(1+x^2)}\,dx.$$
Здесь внутренние интегралы уже сходятся на соответствующих множествах равномерно:
$$\int_0^\infty ye^{-y^2(1+x^2)}\,dx \overset{y\geqslant\delta}{\rightrightarrows} Je^{-y^2}, \qquad \int_\delta^\infty ye^{-y^2(1+x^2)}\,dy \overset{x\geqslant 0}{\rightrightarrows} \Phi(x).$$
Последовательный интеграл тоже сходится, поэтому можно менять порядок интегрирования. Следовательно,
$$J^2 + O(\delta) = \int_0^\infty \frac{1}{2+2x^2}e^{-\delta^2(1+x^2)}\,dx.$$
А теперь перейдём к пределу при $\delta \to 0$, слева получится J^2, а справа интеграл равномерно по δ сходится по критерию Вейерштрасса. Потому можно поменять интеграл и предел при $\delta \to 0$ местами и получится $J^2 = \frac{\pi}{4}$.

Пример 4. Интеграл Дирихле $\displaystyle\int_0^\infty \frac{\sin(xy)}{x}dx = \frac{\pi}{2}\operatorname{sign} y$.

1) Для $y = 0$ очевидно; пусть $y \neq 0$, в силу нечетности достаточно доказать для $y > 0$, замена переменных $t = xy$ приводит к необходимости доказывать только равенство
$$I = \int_0^\infty \frac{\sin x}{x}dx = \frac{\pi}{2}.$$
Положим $\displaystyle F(a,m) = \int_0^\infty \frac{\sin(mx)e^{-ax}}{x}dx$, здесь $m \in [0,1]$, $a \in [0,1]$. Очевидно, $F(0,1) = I$.

2) Рассмотрим интеграл $\displaystyle F'_m(a,m) = \int_0^\infty \cos(mx)e^{-ax}dx$ при $a > 0$. Пользуемся тем, что интегралы для F и F' мажорируются быстро сходящейся экспонентой и по признаку Вейерштрасса равномерно по m сходятся. Воспользуемся равенствами
$$\int_0^\infty \cos(mx)e^{-ax}dx = \int_0^\infty \Re\Big(e^{imx}\Big)e^{-ax}dx = \int_0^\infty \Re\Big(e^{imx}e^{-ax}\Big)dx = \Re\int_0^\infty e^{imx}e^{-ax}dx =$$
$$= \Re\int_0^\infty e^{(im-a)x}dx = \Re\frac{e^{(im-a)x}}{im-a}\Big|_0^\infty = -\Re\frac{1}{im-a} = \Re\frac{im+a}{m^2+a^2} = \frac{a}{m^2+a^2}.$$
То же самое получится, если два раза проинтегрировать по частям:
$$\int_0^\infty \cos(mx)e^{-ax}dx = \frac{1}{a} - \frac{m^2}{a^2}\int_0^\infty \cos(mx)e^{-ax}dx, \qquad F'_m(a,m) = \frac{a}{m^2+a^2}.$$
3) $F(a,m) = \operatorname{arctg}\dfrac{m}{a} + C(a)$, $F(a,0) = 0 \;\Rightarrow\; C(a) = 0 \Rightarrow\; F(a,m) = \operatorname{arctg}\dfrac{m}{a}$.

4) Положим $m = 1$, получим при $a > 0$
$$G(a) = \int_0^\infty \frac{\sin(x)e^{-ax}}{x}dx = \operatorname{arctg}\frac{1}{a}.$$

Хочется перейти к пределу при $a = +0$ и всё. Для этого надо показать, что функция $G(a)$ непрерывна при $a \in [0,1]$ в нуле.

Интеграл от 0 до ∞ разбить на 2 части: $[0,1]$ и $[1,\infty)$. Функция $\displaystyle\int_0^1 \frac{\sin(x)e^{-ax}}{x}dx$ непрерывна — это собственный интеграл. Функция $\displaystyle\int_1^\infty \frac{\sin(x)e^{-ax}}{x}dx$ равномерно сходится по признаку Дирихле: выражение $\displaystyle\int_1^u \sin(x)dx$ равномерно по a, u ограничено, функция $x^{-1}e^{-ax}$ монотонно убывает по x и равномерно по a стремиться к 0. Поэтому несобственный интеграл для G сходится равномерно на $[0,1]$, поэтому предельная функция непрерывна. Переходим к пределу в обеих частях равенства, получаем требуемую формулу. $\square$

Ещё возможность вывода формулы для интеграла Дирихле. Берем равенство

$$\int_0^\infty \frac{\sin(x)}{x}dx = \int_0^\infty \sin x\, dx \int_0^\infty e^{-xy}\, dy = \int_0^\infty dy \int_0^\infty e^{-xy} \sin x\, dx,$$

поменяв в повторном интеграле справа интегралы местами. Но $\displaystyle\int_0^\infty e^{-xy}\sin x\, dx = \frac{1}{1+y^2}$. Поэтому интеграл Дирихле равен $\pi/2$. Осталось доказать, что можно в этом интеграле менять местами пределы интегрирования. Кто будет знать как, тот на экзамене сможет рассказать таким способом. Надо понимать, почему на каждом отрезке есть равномерная сходимость. И вообще, она есть?

Лекция 4 (18 апреля 2016)

На прошлой лекции мы рассмотрели несобственные интегралы, зависящие от параметра.

Были доказаны теоремы о равномерной сходимости таких интегралов, как их дифференцировать, интегрировать, как переставлять несобственные интегралы.

В качестве основного примера посчитали интегралы Дирихле и Эйлера–Пуассона.

Сегодня мы рассмотрим Гамма-функцию и различные формулы для неё, в частности, формулу Стирлинга.

Гамма-функция Рассмотрим вспомогательные функции

$$\Gamma_n(x) = \frac{(n-1)!n^x}{\prod_{k=0}^{n-1}(x+k)}, \qquad n = 2, 3, \ldots \quad x = \mathbb{R} \setminus \{0, -1, -2, -3, \ldots\}$$

Очевидно, что это эквивалентно равенству

$$\Gamma_n(x) = \frac{1 \cdot 2 \cdot \ldots \cdot (n-1)}{x \cdot (x+1) \cdot \ldots \cdot (x+n-1)} \times \frac{n^x}{(n-1)^x} \times \frac{(n-1)^x}{(n-2)^x} \times \ldots \times \frac{2^x}{1^x}.$$

Теперь отдельно рассмотрим оба множителя:

$$\frac{n^x}{(n-1)^x} \times \frac{(n-1)^x}{(n-2)^x} \times \ldots \times \frac{2^x}{1^x} = \prod_{k=1}^{n-1} \left(\frac{k+1}{k}\right)^x = \prod_{k=1}^{n-1} \left(1 + \frac{1}{k}\right)^x$$

и

$$\frac{1 \cdot 2 \cdot \ldots \cdot (n-1)}{x \cdot (x+1) \cdot \ldots \cdot (x+n-1)} = \frac{1}{x} \prod_{k=1}^{n-1} \frac{k}{x+k} = \frac{1}{x} \prod_{k=1}^{n-1} \left(\frac{1}{\frac{x+k}{k}}\right) = \frac{1}{x} \prod_{k=1}^{n-1} \left(\frac{1}{1+\frac{x}{k}}\right),$$

отсюда следует равенство

$$x\Gamma_n(x) = \prod_{k=1}^{n-1} \frac{\left(1+\frac{1}{k}\right)^x}{1+\frac{x}{k}}.$$

Исследуем бесконечное произведение $\Pi = \prod_{k=1}^{\infty} \frac{\left(1+\frac{1}{k}\right)^x}{1+\frac{x}{k}}$.

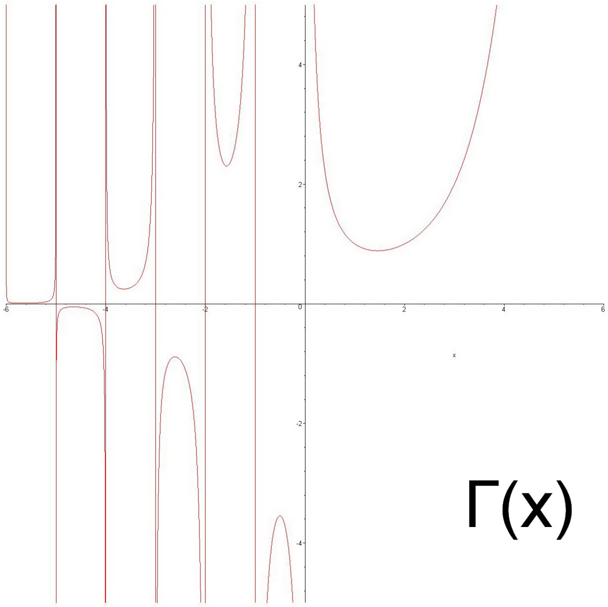
По теореме с одной из предыдущих лекций оно сходится если и только если сходится ряд

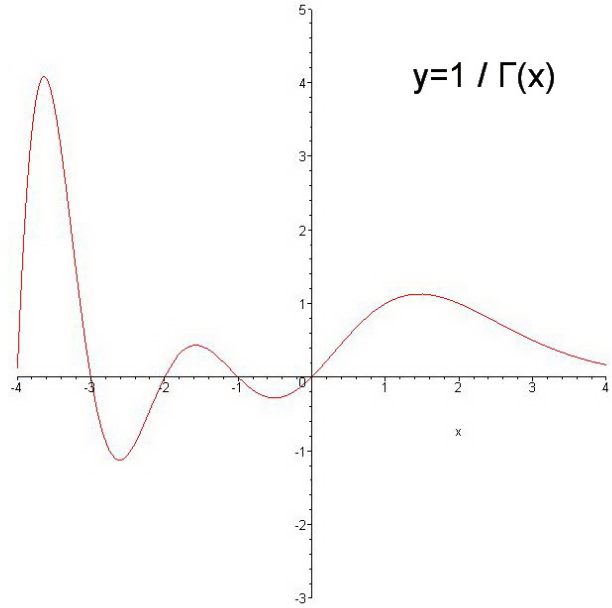
$$\sum_{k=1}^{\infty} \ln \left(\frac{\left(1+\frac{1}{k}\right)^x}{1+\frac{x}{k}}\right) = \sum_{k=1}^{\infty} \left(x \ln\left(1 + \frac{1}{k}\right) - \ln\left(1 + \frac{x}{k}\right)\right) = \sum_{k=1}^{\infty} O(\frac{1}{k^2}).$$

Таким образом произведение Π сходится абсолютно к некоторой функции $\Gamma(x) = \lim\limits_{n\to\infty} \Gamma_n(x)$ и справедливо её представление в виде бесконечного произведения:

$$\Gamma(x) = \frac{1}{x} \prod_{k=1}^{\infty} \frac{\left(1+\frac{1}{k}\right)^x}{1+\frac{x}{k}}.$$

Область определения функции $\Gamma(x)$ — множество $\mathbb{R} \setminus \{0, -1, -2, -3, \ldots\}$.





Основное функциональное равенство $\mathbf{\Gamma(x+1) = x\Gamma(x)}$. Доказательство.

$$\frac{\Gamma(x+1)}{\Gamma(x)} = \frac{\lim_{n\to\infty}\Gamma_n(x+1)}{\lim_{n\to\infty}\Gamma_n(x)} = \lim_{n\to\infty}\frac{\Gamma_n(x+1)}{\Gamma_n(x)} = \lim_{n\to\infty}\frac{\frac{(n-1)!n^{x+1}}{\prod_{k=0}^{n-1}(x+1+k)}}{\frac{(n-1)!n^x}{\prod_{k=0}^{n-1}(x+k)}} = \lim_{n\to\infty}\frac{nx}{n+x} = x \qquad \square$$

Равенство $\mathbf{\Gamma(n+1) = n!}$ Очевидно, что $\Gamma_n(1) = 1 \quad \Rightarrow \quad \Gamma(1) = 1,\ \ \Gamma(2) = 1\cdot\Gamma(1) = 1,$ $\Gamma(3) = 2\cdot\Gamma(2) = 2,\ \ \Gamma(4) = 3\cdot\Gamma(3) = 6,\ \ \ldots\Gamma(n+1) = n!$

4 Эйлеровы интегралы

Теорема. *Справедливо равенство* $\Gamma(x) = \displaystyle\int_0^\infty y^{x-1}e^{-y}dy, \quad x > 0$ *(Первый интеграл Эйлера).*

Доказательство. 1) Докажем, что $\Gamma_{n+1}(x) = (n+1)^x \displaystyle\int_0^1 (1-t)^n t^{x-1}dt$. Для этого по частям:

$$\int_0^1 (1-t)^n t^{x-1}dt = \frac{1}{x}\int_0^1 (1-t)^n d(t^x) = \frac{1}{x}(1-t)^n t^x\Big|_{t=0}^{t=1} + \frac{n}{x}\int_0^1 (1-t)^{n-1}t^x dt = \frac{n}{x}\int_0^1 (1-t)^{n-1}t^x dt =$$

$$= \frac{n(n-1)}{x(x+1)}\int_0^1 (1-t)^{n-2}t^{x+1}dt = \ldots = \frac{n!}{x(x+1)\ldots(x+n-1)}\int_0^1 t^{x+n-1}dt = \frac{n!}{x(x+1)\ldots(x+n)}.$$

Таким образом, требуемое равенство доказано.

2) Положим в этом равенстве $t = y/n$:

$$\Gamma_{n+1}(x) = (n+1)^x \int_0^1 (1-t)^n t^{x-1} dt = (n+1)^x \int_0^n \left(1-\frac{y}{n}\right)^n \frac{y^{x-1}}{n^x} dy = \left(1+\frac{1}{n}\right)^x \int_0^n \left(1-\frac{y}{n}\right)^n y^{x-1} dy.$$

Теперь перейдем к пределу при $n \to \infty$ в последнем равенстве:

$$\lim_{n\to\infty} \Gamma_{n+1}(x) = \Gamma(x), \qquad \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^x = 1, \qquad \lim_{n\to\infty} \int_0^n \left(1-\frac{y}{n}\right)^n y^{x-1} dy = \int_0^\infty e^{-y} y^{x-1} dy. \qquad \square$$

3) Осталось проследить последний предельный переход:

$$\lim_{n\to\infty} \int_0^n \left(1-\frac{y}{n}\right)^n y^{x-1} dy = \int_0^\infty e^{-y} y^{x-1} dy.$$

Для этого оценим модуль разности

$$\int_0^n \left(1-\frac{y}{n}\right)^n y^{x-1} dy - \int_0^n e^{-y} y^{x-1} dy + \int_0^n e^{-y} y^{x-1} dy - \int_0^\infty e^{-y} y^{x-1} dy.$$

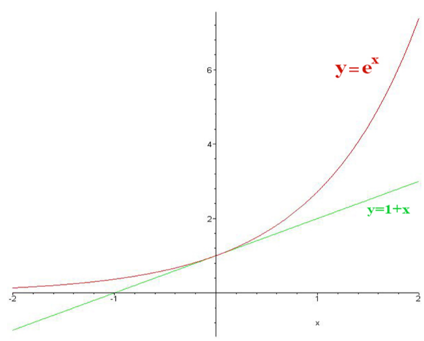
$\left|\int_0^n e^{-y} y^{x-1} dy - \int_0^\infty e^{-y} y^{x-1} dy\right| \to 0$ при $x > 0$ (из сходимости несобственного интеграла),

$$\left|\int_0^n \left(1-\frac{y}{n}\right)^n y^{x-1} dy - \int_0^n e^{-y} y^{x-1} dy\right| = \left|\int_0^n \left[\left(1-\frac{y}{n}\right)^n - e^{-y}\right] y^{x-1} dy\right|$$

так как, $\left(1-\dfrac{y}{n}\right)^n \to e^{-y}$, то $\left|\int_0^1 \left[\left(1-\dfrac{y}{n}\right)^n - e^{-y}\right] y^{x-1} dy\right| \to 0$, теперь осталось оценить

$\left|\int_1^n \left[\left(1-\dfrac{y}{n}\right)^n - e^{-y}\right] y^{x-1} dy\right|$, здесь подинтегральное выражение $\to 0$, промежуток $\to \infty$.

4) Очевидно по картинке, что $e^x \geqslant 1+x$ при всех x. То есть, одновременно, $e^{-x} \geqslant 1 - x$. Положим $x = y/n$, получим $e^{-y} \geqslant \left(1-\dfrac{y}{n}\right)^n$, $\quad e^y \geqslant \left(1+\dfrac{y}{n}\right)^n$. Теперь $e^y\left(1-\dfrac{y}{n}\right)^n \geqslant \left(1+\dfrac{y^2}{n^2}\right)^n$ и, значит, $\left(1-\dfrac{y}{n}\right)^n \geqslant e^{-y}\left(1+\dfrac{y^2}{n^2}\right)^n$.



Теперь

$$0 \leqslant e^{-y} - \left(1-\frac{y}{n}\right)^n \leqslant e^{-y} - e^{-y}\left(1+\frac{y^2}{n^2}\right)^n.$$

Применим к правой части неравенство Бернулли: $(1+s)^n \geqslant 1 + ns$ при $s \geqslant -1$. При $y \leqslant n \Rightarrow -y^2/n^2 \geqslant -1$ и

$$e^{-y} - \left(1-\frac{y}{n}\right)^n \le e^{-y}\,\frac{y^2}{n} \quad \Rightarrow \quad \left|\int_1^n \left[\left(1-\frac{y}{n}\right)^n - e^{-y}\right] y^{x-1} dy\right| \le \frac{1}{n}\left|\int_1^\infty e^{-y} y^{x+1} dy\right| \to 0. \qquad \square$$

Формула дополнения: при $x \in \mathbb{R} \setminus \mathbb{Z}$ справедливо $\Gamma(1-x)\Gamma(x) = \dfrac{\pi}{\sin(\pi x)}$.

Доказательство. Используем формулу разложения синуса в бесконечное произведение

$$x \prod_{k=1}^{\infty} \left(1 - \frac{x^2}{\pi^2 k^2}\right) = \sin(x).$$

Так как $\Gamma(1-x) = -x\Gamma(-x)$, то $\Gamma(1-x)\Gamma(x) = -x\Gamma(-x)\Gamma(x)$, одновременно

$$-x\Gamma(-x)\Gamma(x) = \frac{1}{x} \prod_{k=1}^{\infty} \frac{\left(1+\frac{1}{k}\right)^{-x}}{1-\frac{x}{k}} \prod_{k=1}^{\infty} \frac{\left(1+\frac{1}{k}\right)^{x}}{1+\frac{x}{k}} = \frac{1}{x} \prod_{k=1}^{\infty} \frac{1}{1-\frac{x^2}{k^2}} = \frac{\pi}{\pi x \prod_{k=1}^{\infty}\left(1-\frac{x^2}{k^2}\right)} = \frac{\pi}{\sin(\pi x)}.$$

Интеграл Эйлера-Пуассона $\displaystyle\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$.

Доказательство. Положим $x^2 = t,\ dx = \dfrac{dt}{2\sqrt{t}}$,

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = 2\int_{0}^{+\infty} e^{-x^2} dx = 2\int_{0}^{+\infty} t^{\frac{1}{2}-1} e^{-t} dx = \Gamma(\frac{1}{2}),$$

по формуле дополнения $\Gamma(\dfrac{1}{2})\Gamma(\dfrac{1}{2}) = \pi$. □

Формула удвоения Лежандра: $2^{2a-1}\Gamma(a)\Gamma(a+\dfrac{1}{2}) = \sqrt{\pi}\Gamma(2a)$.

Следует из формулы умножения *Гаусса*

$$\Gamma(a+\frac{1}{m})\Gamma(a+\frac{2}{m})\ldots\Gamma(a+\frac{m-1}{m}) = (2\pi)^{(m-1)/2} m^{-ma+1/2}\Gamma(ma).$$

5 Бета-функция и формула Стирлинга

Положим $B(\alpha,\beta) = \displaystyle\int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt, \quad \alpha,\beta > 0.$

Свойства: 1) Симметричность $B(\alpha,\beta) = B(\beta,\alpha)$ следует из замены $t := 1-t$.

2) Формула понижения: для $\alpha > 1$ верно равенство

$$B(\alpha,\beta) = \frac{\alpha-1}{\alpha+\beta-1} B(\alpha-1,\beta).$$

Для доказательства надо написать формулы:

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = -\frac{1}{\beta} x^{\alpha-1}(1-x)^{\beta}\Big|_0^1 + \frac{\alpha-1}{\beta}\int_0^1 x^{\alpha-2}(1-x)^{\beta} dx =$$

$$= \frac{\alpha-1}{\beta}\int_0^1 x^{\alpha-2}\Big((1-x)^{\beta-1} - x(1-x)^{\beta-1}\Big)dx = \frac{\alpha-1}{\beta}B(\alpha-1,\beta) - \frac{\alpha-1}{\beta}B(\alpha,\beta).$$

3) Докажем связь с Γ–функцией: $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

Это напоминает нам формулу для сочетаний: $C_n^k = \dfrac{(n-k)!k!}{n!}$, хотя и не совсем.

3.1) В силу формулы понижения достаточно доказать формулу для $\alpha > 1$:

$$B(\alpha-1,\beta) = \frac{\alpha+\beta-1}{\alpha-1}B(\alpha,\beta) = \frac{\alpha+\beta-1}{\alpha-1}\,\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \frac{\alpha+\beta-1}{\alpha-1}\,\frac{(\alpha-1)\Gamma(\alpha-1)\Gamma(\beta)}{\Gamma(\alpha+\beta-1)(\alpha+\beta-1)}.$$

Аналогично, достаточно доказать формулу для $\beta > 1$. На самом деле, ещё надо сказать, что для $\alpha = 1$ формула очевидна.

3.2) Замена $t = 1/(1+x),\ 1-t = x/(1+x),\ dt = -t^2dx \Rightarrow B(\alpha,\beta) = \displaystyle\int_0^\infty \frac{x^{\beta-1}}{(1+x)^{\alpha+\beta}}dx$.

3.3) $\Gamma(u) = \displaystyle\int_0^\infty y^{u-1}e^{-y}dy$, пусть $y = (x+1)z$, где x — параметр, а z — новая переменная, тогда $\Gamma(u) = \displaystyle\int_0^\infty (x+1)^u z^{u-1}e^{-(x+1)z}dz$. Эта формула справедлива при всех $x \geqslant 0$.

3.3) $B(\alpha,\beta)\cdot\Gamma(\alpha+\beta) =$

$$= \int_0^\infty \frac{x^{\beta-1}}{(1+x)^{\alpha+\beta}}\,\Gamma(\alpha+\beta)\,dx = \int_0^\infty \frac{x^{\beta-1}}{(1+x)^{\alpha+\beta}}\left(\int_0^\infty (x+1)^{\alpha+\beta}z^{\alpha+\beta-1}e^{-(x+1)z}\,dz\right)dx =$$
$$= \int_0^\infty x^{\beta-1}\left(\int_0^\infty z^{\alpha+\beta-1}e^{-(x+1)z}dz\right)dx \overset{!!!}{=} \int_0^\infty\left(\int_0^\infty x^{\beta-1}z^{\alpha+\beta-1}e^{-(x+1)z}dx\right)dz =$$
$$= \int_0^\infty e^{-z}z^{\alpha-1}\left(\int_0^\infty x^{\beta-1}z^{\beta}e^{-xz}dx\right)dz = \int_0^\infty e^{-z}z^{\alpha-1}\left(\int_0^\infty (xz)^{\beta-1}e^{-(xz)}d(xz)\right)dz = \Gamma(\alpha)\Gamma(\beta). \quad \square$$

Почему можно менять пределы интегрирования по x и по z (там, где !!!)?

Один интеграл сходится равномерно по x признаку Вейерштрасса:

$$\int_0^\infty x^{\beta-1}z^{\alpha+\beta-1}e^{-(x+1)z}dz \rightrightarrows \ldots \qquad \Big(x^{\beta-1}z^{\alpha+\beta-1}e^{-(x+1)z} \leqslant z^\alpha e^{-z}\max(u^{\beta-1}e^{-u})\Big).$$

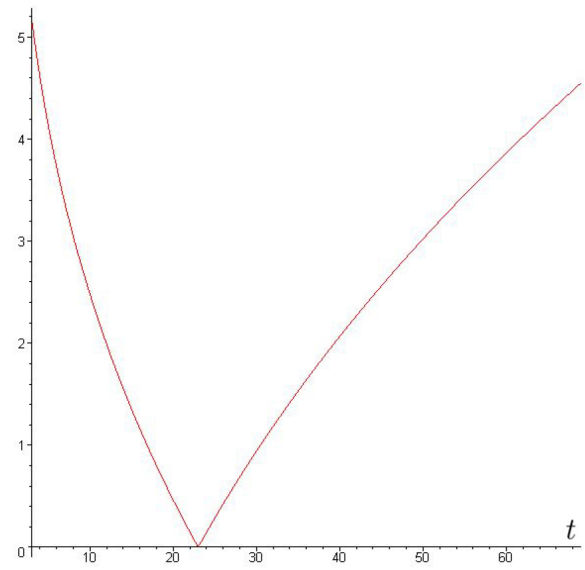
Второй интеграл сходится равномерно по z на каждом конечном промежутке по теореме Дини:

$$\int_0^\infty x^{\beta-1}z^{\alpha+\beta-1}e^{-(x+1)z}dx = z^{\alpha-1}e^{-z}\int_0^\infty (xz)^{\beta-1}e^{-xz}z\,dx \rightrightarrows z^{\alpha-1}e^{-z}\Gamma(\beta),$$

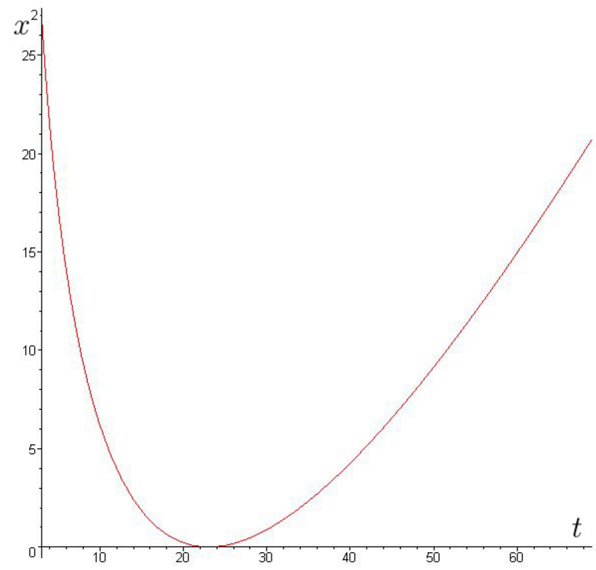
поточечная сходимость к непрерывной функции ($\alpha > 1$), под интегралом — неотрицательная непрерывная функция, повторный интеграл от неё существует.

Пример. Вычислим (замена $t = \sin^2 x,\ x = \arcsin\sqrt{t},\ dx = \dfrac{dt}{2\sqrt{t}\sqrt{1-t}}$)

$$\int_0^{\pi/2}\sin^a x\cos^b x dx = \frac{1}{2}\int_0^1 t^{(a-1)/2}(1-t)^{(b-1)/2}dt = \frac{1}{2}B(\frac{a+1}{2},\frac{b+1}{2}) = \frac{\Gamma(\frac{a+1}{2})\Gamma(\frac{b+1}{2})}{2\Gamma(\frac{a+b}{2}+1)}$$



$x = \sqrt{-(n - t + n\ln t - n\ln n)},\ n = 23$



$x^2 = -(n - t + n\ln t - n\ln n),\ n = 23$

Лекция 5 (19 апреля 2016)

Формула Стирлинга

$n! = (\frac{n}{e})^n(\sqrt{2\pi n} + \alpha), \qquad |\alpha| < 2.$

Вроде бы, это Муавр[1] доказал почти эту формулу, а Стирлинг[2] только написал, что постоянная равна $\sqrt{2\pi}$.

1) $n! = \Gamma(n+1) = \int_0^\infty t^n e^{-t} dt = \left(\frac{n}{e}\right)^n \int_0^\infty \left(\frac{t}{n}\right)^n e^{n-t} dt = \left(\frac{n}{e}\right)^n \int_0^\infty e^{n-t+n\ln t - n\ln n} dt$

2) При $f(t) = n - t + n\ln t - n\ln n$, $f'(t) = -1 + n/t$, поэтому $f'(0) = 0$ только при $t = n$, это точка максимума, $f(t) \geqslant f(n) = 0$, то есть $n - t + n\ln t - n\ln n \leqslant 0$.

3) Делаем замену: $-x^2 = n - t + n\ln t - n\ln n$,

$$x^2 = t - n + n\ln t - n\ln n = t - n - n\ln t + n\ln n = t - n - n\ln\left(1 + \frac{t-n}{n}\right)$$

4) Так как по формуле Тейлора $f(z) = f(0) + f'(0)z + \frac{1}{2}f''(\theta z)z^2$, то

$$\ln(1+z) = z - \frac{z^2}{2}\frac{1}{(1+\theta z)^2}, \qquad \theta = \theta(z)$$

[1] Абрахам де Муавр, 1667-1754, Лондон, формула Муавра, друг Ньютона и Галлея. Труды по теории вероятностей.

[2] Джеймс Стирлинг, 1692-1770. Не надо думать, что он — англичанин. Нет, он шотландец. Отказался присягать английской королеве, уехал в Италию, где и занимался наукой. Изучал алгебраические кривые 3-го порядка, изученные Ньютоном, открыл 4 новых типа, не замеченных Ньютоном.

при каждом z при некотором $\theta \in (0,1)$.

5) Теперь

$$t-n-n\ln\left(1+\frac{t-n}{n}\right)=t-n-n\left(\frac{t-n}{n}-\frac{(t-n)^2}{2n^2}\frac{1}{\left(1+\theta\frac{t-n}{n}\right)^2}\right)=\frac{n(t-n)^2}{2\big(n+\theta(t-n)\big)^2}$$

то есть $x^2=\dfrac{n(t-n)^2}{2\big(n+\theta(t-n)\big)^2}\Rightarrow x=\dfrac{|t-n|\sqrt{\frac{n}{2}}}{n+\theta(t-n)}$ (очевидно, что $n+\theta(t-n)=n(1-\theta)+t\theta\geqslant 0$.)

6) $2xdx=-dt+\dfrac{n}{t}dt\ \Rightarrow\ dt=\dfrac{2xt\,dx}{t-n}$.

7) Теперь «раскроем» модуль в пункте 5. Возможны 2 случая: $t\leqslant n$ и $n\leqslant t$.

8) Пусть $0\leqslant t\leqslant n$. Тогда $x=\dfrac{(n-t)\sqrt{\frac{n}{2}}}{n+\theta(t-n)}\ \Rightarrow\dfrac{2xt}{t-n}=-2\left(\sqrt{\dfrac{n}{2}}-x(1-\theta)\right)$. Куча простых выкладок пропущено, чтобы не пугать читателя. Поэтому

$$\int_0^n e^{n-t+n\ln t-n\ln n}dt=2\int_{-\infty}^0 e^{-x^2}\left(\sqrt{\frac{n}{2}}-x(1-\theta)\right)dx=$$
$$2\left(\sqrt{\frac{n}{2}}\int_{-\infty}^0 e^{-x^2}\,dx+\int_{-\infty}^0 e^{-x^2}x(1-\theta)\,dx\right)=\frac{1}{2}\sqrt{2\pi n}+2\int_{-\infty}^0 e^{-x^2}x(1-\theta)\,dx.$$

8) Аналогично, пусть $t\geqslant n$. Тогда

$$x=\frac{(t-n)\sqrt{\frac{n}{2}}}{n+\theta(t-n)}\ \Rightarrow\frac{2xt}{t-n}=2(\sqrt{\frac{n}{2}}+x(1-\theta)).$$

Поэтому $\displaystyle\int_n^{+\infty} e^{n-t+n\ln t-n\ln n}dt=\frac{1}{2}\sqrt{2\pi n}+2\int_0^\infty e^{-x^2}(1-\theta)x\,dx.$

9) Отсюда

$$\left(\frac{n}{e}\right)^n\int_0^\infty e^{n-t+n\ln t-n\ln n}dt=\left(\frac{n}{e}\right)^n(\sqrt{2n\pi}+2r_n),\qquad r_n=\int_{-\infty}^{+\infty} e^{-x^2}(1-\theta)x\,dx.$$

10) $0\leqslant 1-\theta\leqslant 1\qquad\Rightarrow\qquad |r_n|\leqslant 2\displaystyle\int_0^{+\infty}e^{-x^2}x\,dx=\int_0^\infty e^{-u}\,du=1.$

Формула Стирлинга доказана. □

Более точная формула: $n!=\sqrt{2\pi n}\left(\dfrac{n}{e}\right)^n\left(1+\dfrac{1}{12n}+\dfrac{1}{288n^2}-\dfrac{139}{51840n^3}-\dfrac{571}{2488320n^4}+\ldots\right)$.

Начинаем совсем новую тему.

В первом семестре вам рассказывали про метрические пространства, про пространство $\mathbb{R}^n$, про функции многих переменных, про отображения конечномерных пространств, их непрерывность и их производные. Кое-что вам рассказывали в курсе алгебры.

До конца модуля будем заниматься функциями многих переменных и отображениями в конечномерных пространствах. В основном, функции все будут гладкими, то есть у них будут существовать производные, чаще всего этих производных будет как минимум одна, реже как минимум 2.

На первой лекции я не расскажу вам ничего особенно нового, однако как обзор она будет полезная. Изложение содержание этого модуля в основном следует учебнику Зорича.

6 Метрическое пространство $\mathbb{R}^n$

Что такое пространство $\mathbb{R}^n$, я не буду подробно рассказывать. Линейное вещественное пространство, имеющее размерность n, размерность в обычном смысле — есть n линейно независимых векторов и нет $n+1$ линейно независимых. Считаем, что задано скалярное произведение (x, y) с обычными аксиомами, введем евклидову норму, порождённую этим скалярным произведением: $\|x\| = \sqrt{(x, x)}$, а норма порождает метрику $\|x - y\|$.

То, что аксиомы скалярного произведения (линейность, симметричность и $(x, x) \geqslant 0$) порождают норму, факт трюковый. Положительная однородность очевидна. Для доказательства неравенства треугольника надо рассмотреть выражение $(x + ty, x + ty)$ при вещественных t. Это квадратный трёхчлен относительно t, положительный при всех $t \in \mathbb{R}$. Значит дискриминант у этого трёхчлена — отрицательный: $|(x, y)| \leqslant \|x\| \cdot \|y\|$. Отсюда следует неравенство треугольника.

Базис в $\mathbb{R}^n$ — это любые n линейно независимых векторов, ортонормированный базис — это тот, в котором все базисные векторы ортогональны и имеют длину 1. Про ортогонализацию базиса (как сделать из базиса другой, но ортогональный) — нарисовать картинку для плоскости.

Поэтому справедлива формула для метрики через координаты в ортонормированном базисе:

$$e_k : \ (e_k, e_j) = \delta_j^k, \qquad x = \sum_{k=1}^{n} (x, e_k) e_k. \qquad \Rightarrow \qquad \|x\|^2 = \sum_{k=1}^{n} x_k^2.$$

Окрестности $U_\delta(x_0) = B(x_0, \delta)$ — открытые шары, открытые множества, замкнутые множества. Стандартное определение предельных точек, внутренних точек, граничных точек множеств.

Последовательности в $\mathbb{R}^n$. Пределы, сходимость по евклидовой норме эквивалентна покоординатной сходимости. И всякой сходимости по любой норме, вам это на алгебре рассказывали. Я чуть позже ещё раз расскажу.

Полнота $\mathbb{R}^n$. Всякая фундаментальная последовательность в $\mathbb{R}^n$ сходится. В самом деле, пусть задана фундаментальная последовательность $x^1, x^2 \ldots \in \mathbb{R}^n$. Тогда последовательность чисел (x^n, e_k) проекций x^n на фиксированный базисный вектор e_k тоже фундаментальная. Поэтому все последовательности (x^n, e_k) (при каждом k от 1 до n) сходятся, к числам y_k. Теперь, очевидно, $x^n \to \sum y_k e_k$.

Другой вариант полноты: вложенные кубики сходятся к точке — это очевидное покоординатное следствие из теоремы в $\mathbb{R}$.

На самом деле, последовательность вложенных замкнутых подмножеств с диаметрами, стремящимися к 0, имеет общую точку, причём единственную.

Компактность замкнутого куба в $\mathbb{R}^n$. Компактность — если из любого открытого покрытия можно извлечь конечное подпокрытие.

Теперь докажем компактность куба «делением отрезка пополам». Отсюда следует компактность любого ограниченного замкнутого множества в $\mathbb{R}^n$ (добавим ещё одно открытое множество — дополнение до открытого куба чуть большего размера). Секвенциальная компактность: это то, что из каждой ограниченной последовательности можно извлечь сходящуюся. Извлечем подпоследовательность сначала так, чтобы первая координата сходилась, потом вторая. В результате получится сходящаяся подпоследовательность.

Непрерывные функции и отображения на компактах. Определение непрерывной функции и непрерывного отображения.

Свойства непрерывных функций (локальные: элементарные операции, сложная функция; на компакте: ограниченность, минимум и максимума, теорема Кантора). Непрерывность отображений покоординатная.

Единственный хитрый вопрос (которого не было в $\mathbb{R}$): есть функция $f(x, y)$. Когда

$$\lim_{y\to y_0} \lim_{x\to x_0} f(x,y) = \lim_{x\to x_0} \lim_{y\to y_0} f(x,y) = \lim_{(x,y)\to(x_0,y_0)} f(x,y)?$$

Теорема. *Если функция непрерывна в окрестности точки (x_0, y_0), то повторные пределы совпадают с $f(x_0, y_0)$.*

Доказательство очевидное:

$$\lim_{y\to y_0} \lim_{x\to x_0} f(x,y) = \lim_{y\to y_0} f(x_0,y) = f(x_0,y_0), \qquad \lim_{x\to x_0} \lim_{y\to y_0} f(x,y) = \lim_{x\to x_0} f(x,y_0) = f(x_0,y_0).$$

Если функция f непрерывна всего лишь в точке, то все может быть плохо, повторные пределы могут не существовать.

Пример 1. $f(x, y) = xy/(x^2 + y^2)$ в окрестности нуля, 0 в нуле. Здесь повторные пределы равны 0, а предел по прямой $x = y$ равен $1/2$.

Пример 2. $f(x, y) = (x^2 - y^2)/(x^2 + y^2)$ в окрестности нуля, 0 в нуле. Здесь повторные пределы равны ± 1.

Пример 3. $f(x, y) = x + y \sin(x^{-1})$ при $x \neq 0$, и 0 при $x = 0$. Здесь есть двойной предел, есть один из повторных, а второго нет.

Объяснение: повторные пределы — это пределы по вертикальным и горизонтальным линиям. А нужны пределы «как угодно», они гарантируются непрерывностью в окрестности.

Другие метрики в $\mathbb{R}^n$. Покоординатная сходимость. Пусть в пространстве X есть две нормы, $\|x\|^*$ и $\|x\|^{**}$. Они называются эквивалентными, если

$$\exists c, C > 0 : \quad \forall x \in X \qquad \text{справедливо } \ c\|x\|^* \leqslant \|x\|^{**} \leqslant C\|x\|^*.$$

Эквивалентность норм, в частности, означает, что последовательности, сходящиеся по одной норме, сходятся по другой.

Теорема. *В пространстве $\mathbb{R}^n$ все нормы эквивалентны.*

Эта теорему вам рассказывали и на алгебре, и на топологии. И я повторю.

Прежде чем доказывать, контрпример в C. Нормы в $C(0, 1)$, взятые из L^1 или L^2,

$$\|x\|_{L^1} = \int_0^1 |x(t)|\, dt, \qquad \|x\|_{L^2} = \sqrt{\int_0^1 x^2(t)\, dt},$$

не эквивалентны обычной норме в C: последовательность t^n не сходится по норме в C, но сходится к нулю по интегральным нормам.

Множество непрерывных функций не является полным по обеим интегральным нормам. Вообще, из того, что все нормы в пространстве эквивалентны следует, что пространство конечномерно. Другое похожее утверждение: если замкнутый единичный шар компактен, то пространство конечномерно.

Доказательство. 1. Докажем, что любая норма эквивалентна евклидовой.

2. В одну сторону: $\|x\|^* = \|\sum_{k=1}^{n} x_k e_k\|^* \leqslant \sum_{k=1}^{n} \|x_k e_k\|^* = \sum_{k=1}^{n} x_k \|e_k\|^* \leqslant \|x\| \sqrt{\sum (\|e_k\|^*)^2} = C\|x\|$.

3. Теперь мы знаем, что функция $f(x) = \|x\|^*$ непрерывна:

$$|f(x) - f(x_0)| = \Big| \|x\|^* - \|x_0\|^* \Big| \leqslant \|x - x_0\|^* \leqslant C\|x - x_0\|$$

(в середине использовали неравенство треугольника $\big| \|x\|^* - \|y\|^* \big| \leqslant \|x - y\|^*$ для нормы $\|\cdot\|^*$).

4. Теперь возьмём единичную сферу $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$. Это ограниченное и замкнутое подмножество $\mathbb{R}^n$. Следовательно, это компакт.

5. Функция $f(x) = \|x\|^* > 0$ и непрерывна на компакте S^{n-1}, следовательно, $c = \min f > 0$. Итак, $\|x\|^* = \|x\| \|\frac{x}{\|x\|}\|^* \geqslant c\|x\|$. □

Лекция 6 (25 апреля 2016)

На прошлой лекции мы рассмотрели формулу Стирлинга, в одном из вариантов. И перешли к новой теме.

Я большую часть лекции рассказывал о том, что вы знаете, о конечномерных пространствах, о базисах, нормах, полноте и компактности в R^n.

Линейное отображение. Определение. Его матрица, матрица привязана к базису. Запись Ax, почему не $A(x)$.

Теорема. *Линейное отображение* $A:\mathbb{R}^m\to\mathbb{R}^n$ *ограничено* $\exists C:\|Ax\|_{\mathbb{R}^n}\leqslant C\|x\|_{\mathbb{R}^m}$ *и непрерывно.*

Доказательство: $\|A\sum\limits_{k=1}^m \xi_k e_k\|_{\mathbb{R}^n}=\sum\limits_{k=1}^m|\xi_k|\;\|Ae_k\|_{\mathbb{R}^n}\leqslant\|x\|_{\mathbb{R}^m}\sqrt{\sum\limits_{k=1}^m\left(\|Ae_k\|_{\mathbb{R}^n}\right)^2}.$ □

Липшицевость и непрерывность следует из ограниченности.

Обсудить зависимость–независимость он нормы.

Линейное пространство $\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)$ линейных преобразований из $\mathbb{R}^m$ в $\mathbb{R}^n$. Норма в $\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)$:

$$\|A\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)}=\max_{x\in\mathbb{R}^m,\|x\|_{\mathbb{R}^n}=1}\|Ax\|_{\mathbb{R}^n}=\max_{x\in\mathbb{R}^m,x\neq0}\frac{\|Ax\|_{\mathbb{R}^n}}{\|x\|_{\mathbb{R}^m}}.$$

Поговорить, почему это норма.

Линейный функционал. Общий вид линейного функционала в $\mathbb{R}^m$. Линейные отображения из $\mathbb{R}^1$ — умножение константы из $\mathbb{R}$ на некоторый вектор.

Пространство линейных функционалов — сопряжённое пространство. Норма в сопряжённом пространстве индуцируется нормой в исходном пространстве:

$$\ell(x)=\ell(\sum x_ke_k)=\sum x_k\ell(e_k)=(x,a),\quad a=(\ell(e_1),\ldots,\ell(e_m));\qquad \|\ell\|=\sum_{\|x\|=1}\ell(x).$$

Пример: евклидова норма (которая порождена скалярным произведением) порождает точно такую же норму:

$$|(x,a)|\leqslant\|x\|\cdot\|a\|,|(a,a)|=\|a\|\cdot\|a\|\;\Rightarrow\;\|\ell\|=\|a\|.$$

Какую норму порождает норма $\sum|x_k|$?

$$\sup_{\sum|x_k|=1}|(x,a)|=\sup_{\sum|x_k|=1}|\sum x_ka_k|=\max|a_k|,\Rightarrow\;\|\ell\|=\max|a_k|.$$

Вопросы: какую норму порождает норма $\max|x_k|$? Норма $(\sum|x_k|^p)^{1/p}$?

Дифференцирование отображений в конечномерных пространствах. Рассмотрим пространства $\mathbb{R}^m$ и $\mathbb{R}^n$ с некоторыми выбранными нормами. Так как это конечномерные пространства, то там все нормы эквивалентны: $c\|x\|_1 \leqslant \|x\|_2 \leqslant C\|x\|_1$.

Отображение $F : \mathbb{R}^m \to \mathbb{R}^n$. **Определение дифференцируемости** отображения в точке x: существует линейное отображение (я иногда буду говорить «*оператор*», это то же самое, что «*отображение*», в моих устах — полная тавтология) $A : \mathbb{R}^m \to \mathbb{R}^n$, такое что $F(x+h) - F(x) - Ah = o(\|h\|)$, то есть

$$\lim_{\|h\|\to 0} \|h\|^{-1} \|F(x+h) - F(x) - Ah\| = 0$$

или, что то же,

$$\forall \varepsilon > 0 \ \exists \delta > 0 \ \forall h \in \mathbb{R}^m, \ \|h\| \leqslant \delta \ \text{ справедливо } \ \|h\|^{-1} \|F(x+h) - F(x) - Ah\| \leqslant \epsilon.$$

В определении есть несколько мест, в которых использован значок норму. В конечномерном пространстве все нормы эквивалентны, поэтому производная не зависит от нормы.

Линейное отображение A называют по разному (дифференциал, касательно отображение, производная) и обозначают по-разному: $dF(x), DF(x), F'(x)$. Каждое такое обозначение — единый символ линейного оператора, он такой многобуквенный, для того, чтобы подчеркнуть, что это производная, от какой функции, в какой точке мы считаем эту производную. Обозначать все это принято так: $dF(x) \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ — пространство линейных операторов из $\mathbb{R}^m$ в $\mathbb{R}^n$. Значения производной — это векторы $dF(x)h, DF(x)h, F'(x)h$, h — это приращение аргумента, $F(x+h) - F(x)$ — приращение функции.

Сказать про оператор $x \mapsto \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$, ставящий в соответствие каждой точке x линейный оператор — дифференциал функции f в точке x.

Когда мы говорим про обозначение $\partial F(x)h$ и называем «ЭТО» основной линейной частью приращения $F(x+h) - F(x)$ функции F, то мы сталкиваемся с обычной математической проблемой совпадения обозначений функции и значений функции. Пример: $\sin x$ — это и значение функции синус в точке x, и сама функция $\sin(\cdot)$. Так и тут. Дифференциалом во многих учебниках называют и значение линейного отображения, и само это линейное отображение.

Заметим, что в простейшем случае, производной скалярной функции мы называли число. С точки зрения данного определения, это будет не число, а линейная операция умножения на это число. Ясно, что это формально разные вещи, но по сути — одинаковые.

О других частных случаях — когда $n = 1$ или $m = 1$ я ещё буду говорить.

Линейный оператор — дифференциал $f'(x_0)$ в точке x_0 — действует в пространстве приращений к точке x_0. И значения этот линейный оператор принимает в множестве приращений, причем начало координат переходит в начало координат.

Вот эти самые приращения образуют важную структуру — касательное пространство.

Теперь я поговорю о важном понятии *касательное пространство*. Мы применяем линейный оператор дифференциал всегда к приращениям аргумента. Никогда не к x, всегда к h. Геометрически это означает, что вместо исходного пространства векторов x («торчащих из начала координат») мы рассматриваем множество векторов h, торчащих из x. Его тоже можно рассматривать, как линейное пространство той же размерности. В $\mathbb{R}^m$ и в $T_x\mathbb{R}^m$ разные базисы.

Называется касательное пространство, обозначается $T_x\mathbb{R}^m$, $T\mathbb{R}^m(x)$, $T\mathbb{R}^n_x$. Буква T от слова *tangent* (*касательная*). Значение дифференциала $dF(x)$ на векторе $h \in T_x\mathbb{R}^m$ есть вектор $dF(x)h \in T_{F(x)}\mathbb{R}^n$ — главная линейная часть приращения отображения при малых h.

Не путаем касательное пространство с касательной прямой-плоскостью-гиперплоскостью! Касательная плоскость — это чисто геометрический объект.

Частные производные. До сих пор в определении дифференцируемости у нас были пространства $\mathbb{R}^n$ и $\mathbb{R}^m$ без базисов. Предположим, что мы ввели ортонормированные базисы в обоих этих пространствах, то есть отображение F имеет координатную форму $F(x) = (F^1(x), \dots, F^n(x))$ или $F^1(x_1, \dots, x_m), \dots, F^n(x_1, \dots, x_m)$, где $x = (x_1, \dots, x_m)$.

Утверждение. *Отображение F дифференцируемо в точке x, если и только если дифференцируемы все n скалярнозначных функций F^k.*

Доказательство я приводить не буду, вам его рассказывали, да оно и очевидное.

Теперь мы немножко поговорим о функциях со скалярными значениями. Об их производных, о возникающих при этом понятиях.

Пусть дана функция $f(x_1, \dots, x_m) \mapsto \mathbb{R}$.

Определение. *Частной производной f'_{x_k}* функции f по компоненте x_k называется предел

$$\lim_{h \to 0} h^{-1} \Big(f(x_1, \dots, x_k + h, \dots, x_m) - f(x_1, \dots, x_k, \dots, x_m) \Big),$$

если он существует.

Утверждение. *Если функция дифференцируемая в точке x, то существуют частные производные в точке x. Если функция имеет частные производные в окрестности точки x, и они непрерывны в x, то она дифференцируема в x.*

Это 2 теоремы, одна — о существовании частных производных у дифференцируемой функции, другая — о дифференцируемости функции при наличии непрерывных частных производных, вторая — несколько сложнее.

Соответственно, если частные производные непрерывны в области, то функция дифференцируема в области. Хорошие контрпримеры вы знаете (когда функция непрерывна, когда существуют частные производные в окрестности, но не существует дифференциала).

По определению производная функции — это линейный функционал (линейный функционал на векторном пространстве — это линейная функция на этом пространстве, принимающая скалярные значения). Всякий линейный функционал $\ell(x) \mapsto \mathbb{R}$ в конечномерном пространстве со скалярным произведением порождается некоторым вектором: $\ell(x) = (b, x)$.

Это простое утверждение. Вот есть базис e_k, Рассмотрим b вектор с координатами $\ell(e_k)$. Теперь для любого вектора $x = \sum x_k e_k$ значение функционала имеет вид

$$\ell(x) = \ell(\sum x_k e_k) = \sum \ell(x_k e_k) = \sum x_k \ell(e_k) = (x, b).$$

Иными словами, для всякой дифференцируемой в точке x функции f ее дифференциал имеет вид $df(x)h = (b, h)$, равенство справедливо при всех h. Вот этот вектор b и называется градиентом функции f (в точке x), обозначается $\mathbf{grad}\, f(x)$. Легко посчитать, что вектор $\mathbf{grad}\, f(x)$ имеет компоненты $\partial f/\partial x_k$, он может быть записан в виде $\nabla f(x)$, где ∇ — это символ, обозначающий «вектор»:

$$\nabla = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_m}).$$

Это не вектор, не элемент пространства — воспринимаем просто как обозначение, символ. Чисто мнемонически запоминаем формулу $\mathbf{grad}\, f = \nabla f$, как произведение вектора ∇ на скаляр f. Можно подразумевать, что ∇ — это такой дифференциальный оператор, который определен на гладких функциях f (нужной размерности) и переводящей функцию в её градиент $\mathbf{grad}\, f$.

Производная произведения и частного:

$$(fg)' = f'g + fg', \ \left(\frac{f}{g}\right)' = \frac{1}{g^2}(gf' - fg') \quad \Leftrightarrow \quad \nabla(fg) = g\nabla f + f\nabla g, \ \nabla\left(\frac{f}{g}\right) = \frac{g\nabla f - f\nabla g}{g^2}.$$

Эти формулы были справедливы для обычных вещественных функций, они так же справедливы и для скалярных функций многих переменных, условие — дифференцируемость функций f и g и ненулевое значение g для частного. Здесь в формулах всюду f' и g' — это векторы, они умножаются и делятся на скаляры f и g.

Другой важный частный случай — это функции из $\mathbb{R}$ в $\mathbb{R}^n$. Для таких функций производная определяется проще всего.

Матрица Якоби. Якобиан. Называют якобианом и матрицу, и определитель (чаще). Матрица может быть и не квадратная, якобиан бывает только у квадратных матриц.

Производная — это линейный оператор из $\mathbb{R}^m$ в $\mathbb{R}^n$. Если в обоих этих пространствах выбраны базис, то каждый линейный оператор — это оператор умножения на $m \times n$ матрицу.

Как мы знаем, элементы этой матрицы задаются частными производными $\partial f/\partial x_k$:

$$\begin{pmatrix} \partial f_1/\partial x_1 & \ldots & \partial f_1/\partial x_m \\ \ldots & \ldots & \ldots \\ \partial f_n/\partial x_1 & \ldots & \partial f_n/\partial x_m \end{pmatrix}.$$

Лекция 7 (10 мая 2016)

На прошлой лекции мы рассмотрели определение дифференцируемости функций в конечномерных пространствах.

Производная по вектору и по направлению. Производная по вектору:

$$D_v f(x) := \lim_{t\to 0} \frac{f(x+vt)-f(x)}{t} = (v, \mathbf{grad}\, f).$$

Производная по направлению — производная по вектору $e = (\cos\alpha_1, \ldots, \cos\alpha_m)$ единичной длины:

$$D_e f(x) = (\mathbf{grad}\, f, e) = \sum_{k=1}^{m} \frac{\partial f}{\partial x_k} \cos\alpha_k;$$

α_k — направляющие косинусы, углы α_k — это углы между e и базисными векторами e_k.

Частная производная — производная по направлению координатного вектора.

Характеристическое свойство градиента. *Направление градиента — направление наибольшей скорости функции.*

Дифференцирование композиции отображений.

Основной смысл теоремы: производная композиции равна композиции производных.

Теорема. *Пусть $f : X \to Y$, $g : Y \to Z$, причем отображение f — дифференцируемое в точке x, а g — дифференцируемое в точке $f(x)$, тогда композиция $g(f(x))$ дифференцируемое в точке x и $(g\circ f)' = g' \circ f'$.*

Доказательство. Очевидно, что

$$g(f(x+h)) - g(f(x) = g'(f(x))(f(x+h)-f(x)) + o(f(x+h)-f(x)) =$$
$$= g'(f(x))f'(x)h + o(f(x+h)-f(x)) + g'(f(x))o(h).$$

Теперь $g'(f(x))o(h) = o(h)$ в силу ограниченности каждого конечномерного оператора в конечномерном пространстве и $o(f(x+h)-f(x)) = o(f'(x)h + o(h)) = o(h)$. □

Это общий факт, он справедлив для производных операторов в любых линейных нормированных пространств. Если все расписывать в координатном виде, то получится произведение матриц Якобы, матрицы как раз окажутся нужного размера (они не квадратные), чтобы их можно было умножать.

Старшие производные. Для функций одной переменной мы знаем, что бывает первая производная, а бывают всякие следующие производные: вторая, третья...

Что является естественным аналогом старших производных для вектор функций?

Рассказать про отображение $\mathbb{R}^m \to \mathbb{R}^{mn}$.

Рассказать про полилинейные формы на примере второй производной.

Дать определение старших частных производных и определение классов C^p.

Дифференцирование обратного отображения.

Эта теорема похожа на теорему о неявной функции, про которую я буду много говорить на следующем занятии. Теорема о неявной функции — это теорема о локальной однозначной разрешимости уравнения $F(x, y) = 0$. Там было примерно так: если $F(x_0, y_0) = 0$ и $F'_y(x_0, y_0) \neq 0$, то (при выполнении некоторых дополнительных существенных технических условий) локально $y = f(x)$.

Для обратной функции все почти так же: уравнение $y - f(x) = 0$, векторы x и y одинаковой размерности, Если производная по y существует и обратима в некоторой точке, то существует ли обратный оператор? Не обязательно даже для функций одного переменного! Пример у вас был когда-то в листке:

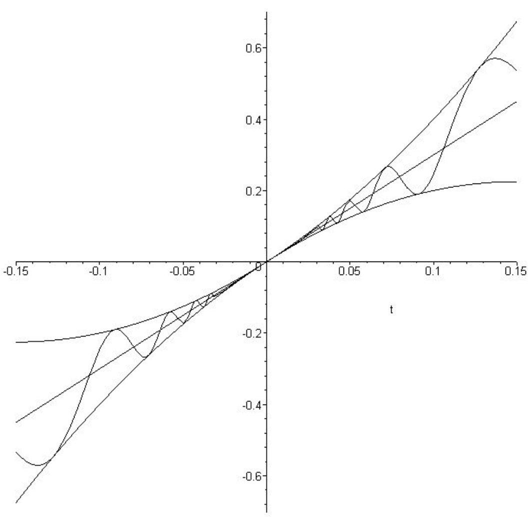


Рис. 1
Это график и функции $y = 3x + 10x^2 \sin(x^{-1})$, и вспомогательных функций $y = 3x \pm 10x^2$ и $y = 3x$. В нуле у функции есть производная, она равна 3, однако в окрестности нуля функция колеблется и у неё обратной функции. А обратная функция есть только у монотонной функции.

Теорема. *Пусть f — отображение окрестности $U(x) \subset \mathbb{R}^m$ точки x на окрестность $V(y) \subset \mathbb{R}^m$ точки y. Пусть f непрерывно в окрестности точки x и имеет обратное отображение $f^{-1} : V(y) \to U(x)$, непрерывное в точке $y = f(x)$.*

Если при этом f дифференцируемо в точке x и производная $f' : T\mathbb{R}^m_x \to T\mathbb{R}^m_y$ обратима: существует $(f'(x))^{-1} : T\mathbb{R}^m_y \to T\mathbb{R}^n_y$, то отображение f^{-1} дифференцируемо в точке y и $(f^{-1})'(y) = (f'(x))^{-1}$.

Теорема о неявной функции. *Пусть задано отображение $F : U \subset \mathbb{R}^{m+n} \to \mathbb{R}^n$, U — окрестность точки (x_0, y_0). Пусть*

1) $F(x_0, y_0) = 0$;

2) $F \in C^p(U, \mathbb{R}^n)$, $p \geqslant 1$;

3) Дифференциал $F'_y(x_0, y_0)$ — обратимый линейный оператор;

Тогда существуют промежутки $I_x^m = \{x \in \mathbb{R}^m : \|x - x_0\| \leqslant \alpha\}$ и $I_y^n = \{y \in \mathbb{R}^n : \|y - y_0\| \leqslant \beta\}$ и такое отображение $f \in C^p(I_x^m, I_y^n)$, что что для любой точки $(x, y) \in I_x^m \times I_y^n$ справедливо равенство

$$F(x,y) = 0 \quad \Leftrightarrow y = f(x), \qquad f'(x) = -\big(F'_y(x, f(x))\big)^{-1} F'_x(x, f(x)).$$

Почти эту теорему мы ещё раз будем доказывать на одной из двух последующих лекций. Последняя формула означает, что в некоторой окрестности точки (x_0, y_0) неявная функция определена, причём однозначным образом. Я через неделю докажу один из вариантов этой теоремы.

Теоремы о конечных приращениях.

Теорема о среднем для скалярнозначных функций. *Пусть отображение f действует из области $D \subset \mathbb{R}^n$ в $\mathbb{R}$. Пусть $h \neq 0$ и $x, x+h \in D$ вместе с отрезком $[x, x+h]$. Пусть f дифференцируемо во всех точках $[x, x+h]$, тогда при некотором $\theta \in [0,1]$ справедливо равенство*

$$f(x+h) - f(x) = df(x + \theta h)h.$$

Равенство можно переписать в виде $f(x+h) - f(x) = (\nabla f(x + \theta h), h)$.

Доказательство. Рассмотрим скалярную функцию $F(t) = f(x + t\,h) : [0,1] \to \mathbb{R}$. Эта функция дифференцируема, для неё справедлива формула Лагранжа конечных приращений: $F(1) - F(0) = F'(\theta)$. Это равенство совпадает с требуемым в теореме. $\square$

Отсюда вытекает важное **следствие.** *Если в линейно связной области D дифференциал функции равен нулю, то функция постоянная.*

Для выпуклой (или звёздчатой) области этот факт очевиден, в любой паре точек функция принимает одинаковые значения, отрезок целиком лежит в области. Для невыпуклых областей ситуация сложнее.

Берем 2 произвольные точки в области. Пусть $x(t), t \in [0,1]$ — путь содержащийся в области, имеющий концами эти 2 точки. Образ компакта при непрерывном отображении — компакт. Покроем каждую точку этого образа–компакта шариком, целиком лежащим в области. Выберем конечное подпокрытие. В каждом шарике по следствию функция постоянна,

Легко видеть, отсюда следует, что функция постоянна везде. Красим точку $x(0)$ в красный цвет. На следующем шаге закрашиваем целиком каждый шарик, имеющий хоть одну красную точку. Во всех красных точках функция имеет одинаковое значение. Очевидно, что в результате конечного числа шагов процесс окраски стабилизируется. Если будут окрашены все шары, то все доказано.

Пусть в результате часть шаров будет закрашена, а часть — нет. Это противоречит замкнутости и связности пути и открытости шаров (берем первую неокрашенную точку, в любой ее окрестности есть окрашенные). $\square$

Лекция 8 (13 мая 2016)

На прошлой лекции мы повторили некоторые вещи, связанные с производными отображений конечномерных пространств, я сформулировал теорему о неявной функции и приступил к её доказательству. В доказательстве используются некоторые понятия и факты. который все равно надо знать.

Начал я с простейшей теоремы о конечных приращениях (формула. похожая на Лагранжа: $f(x+h)-f(x)=df(x+\theta h)h$), продолжу чуть более сложной теоремой о среднем для векторнозначных функций.

Во-первых, ещё раз, мы знаем, что точно такая же теорема не верна. Берем функцию $f:\mathbb{R}\to\mathbb{R}^2$, определяемую равенствами $t\mapsto f(t)=(\cos t,\sin t)$, видим, что для неё ничего не получится: $|f'(t)|=1, f(0)=f(\pi)$.

Однако, важная часть теоремы о среднем может быть сформулирована и для общего случая.

Сначала напомню про **нормы линейных операторов**.

1. Линейные операторы из $\mathbb{R}^m$ в $\mathbb{R}^n$ образуют линейное пространство размерности mn, обозначается $\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)$. В это пространстве можно вводить различные нормы, они все эквивалентны.

2. Примеры: если в пространствах $\mathbb{R}^m$ и $\mathbb{R}^n$ введены базисы, то элементы $\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)$ можно отождествить с матрицами, можно строить всякие нормы по элементам матриц. Например, корень из суммы квадратов, сумма модулей, максимальный модуль элемента. При этом утрачивается основная сущность пространства $\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)$ — то, что это пространство линейных операторов.

3. Повторяю самую главную конструкцию. Пусть в пространствах $\mathbb{R}^m$ и $\mathbb{R}^n$ введены нормы, пока что не важно какие. Тогда назовем нормой $A\in\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)$ величину

$$\|A\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)}=\sup_{e\in\mathbb{R}^m,\|e\|_{\mathbb{R}^m}=1}\|Ae\|_{\mathbb{R}^n}.$$

Это норма, проверить все свойства нормы (только в начале координат норма 0, неравенство треугольника и линейность) легко.

Как мы уже говорили, у каждого отображения конечномерных пространств такая норма определена, она зависит от отображения и норм в пространствах.

Основное определение означает, что для всех $x\in\mathbb{R}^m$ справедливо неравенство $\|Ax\|\leqslant\|A\|\cdot\|x\|$, в котором все нормы имеют естественный смысл: $\|Ax\|_{\mathbb{R}^n}\leqslant\|A\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)}\cdot\|x\|_{\mathbb{R}^m}$.

4. В частности, дифференциал — это тоже линейный оператор. Значит у него тоже есть норма.

5. Замечу, что эта конструкция определения нормы годится для любых линейных нормированных пространств. В бесконечномерных пространствах бывают линейные ограниченные операторы, а бывают и неограниченные.

Как я уже говорил, ограниченность оператора в нормированных пространствах эквивалентна его непрерывности. Непрерывность линейного оператора в одной точке эквивалентна непрерывности в каждой точке и эквивалентна равномерной непрерывности.

6. Пример неограниченного оператора: оператор дифференцирования в пространстве непрерывно дифференцируемых функций на $[0, 1]$ с нормой в C. Тогда функции $\sin(nx)$ лежат на единичном шаре, а норма образа равна n, то есть не ограничена. Конечно, это уродливое пространство, не полное. Если оператор определен на всем пространстве, аддитивен и однороден, то он довольно часто оказывается ограниченным. Например, если оператор действует на всем гильбертовом пространстве (не буду останавливаться пока на точном определении, приблизительно это линейное пространство со скалярным произведением), аддитивен, однороден и симметричен, то он ограничен.

Теорема о среднем. *Пусть отображение f действует из области $D \subset \mathbb{R}^m$ в $\mathbb{R}^n$. Пусть $h \neq 0$ и $x, x+h \in D$ вместе с отрезком $[x, x+h]$. Пусть f дифференцируемо во всех точках $[x, x+h]$ и пусть конечна величина*

$$S = \sup_{\theta\in[0,1]} \|df(x+\theta h)\|_{\mathcal{L}_{\mathbb{R}^m\to\mathbb{R}^n}}.$$

Тогда $\|f(x+h) - f(x)\|_{\mathbb{R}^n} \leqslant S\,\|h\|_{\mathbb{R}^m}$.

Доказательство. Пусть это не так. Тогда при некотором $\varepsilon > 0$ справедливо неравенство $\|f(x+h) - f(x)\|_{\mathbb{R}^n} > (S+\varepsilon)\,\|h\|_{\mathbb{R}^m}$. Разобьем отрезок $[x, x+h]$ пополам, хотя бы на одном из отрезков $[x, x+h/2]$ или $[x+h/2, x+h]$ справедливо неравенство

$$\|f(x+h) - f(x+h/2)\|_{\mathbb{R}^n} > (S+\varepsilon)\,\|h/2\|_{\mathbb{R}^m}, \qquad \|f(x+h/2) - f(x)\|_{\mathbb{R}^n} > (S+\varepsilon)\,\|h/2\|_{\mathbb{R}^m}.$$

Разобьем такой промежуток на 2 части, снова на одном из частей будет выполнено аналогичное неравенство, продолжим эту процедуру далее, получим последовательность вложенных промежутков $[a_k, b_k] \subset [0, 1]$, для которых при всех k справедливы неравенства

$$\|f(x+a_k h) - f(x+b_k h)\|_{\mathbb{R}^n} > (S+\varepsilon)\,|a_k - b_k|\,\|h\|_{\mathbb{R}^m}.$$

Эти отрезки имеют общую точку τ,

Более того, для каждого отрезка $[a_k, b_k]$, для которого $\tau \in (a_k, b_k)$ справедливо аналогичное неравенство, либо на отрезке $[\tau, b_k]$, либо на отрезке $[a_k, \tau]$. Теперь хотя бы одно из множеств таких невырожденных отрезков бесконечно, пусть, например, бесконечно множество стягивающих отрезков $[\tau, b_k]$:

$$\|f(x+\tau h) - f(x+b_k h)\|_{\mathbb{R}^n} > (S+\varepsilon)\,|b_k - \tau|\,\|h\|_{\mathbb{R}^m}.$$

А это неравенство противоречит неравенству $\|df(x+\tau h)\|_{\mathcal{L}_{\mathbb{R}^m\to\mathbb{R}^n}} \leqslant S$, так как

$$f(x+\tau h) - f(x+b_k h) = df(x+\tau h)h(b_k - \tau) + o(|b_k - \tau|). \qquad \square$$

Следствие. *Пусть отображение f действует из замкнутой* ***выпуклой*** *области $D \subset \mathbb{R}^m$ в $\mathbb{R}^n$. Пусть f дифференцируемо во всех точках D и пусть существует*

$$S = \sup_{x \in D} \|df(x)\|_{\mathcal{L}_{\mathbb{R}^m \to \mathbb{R}^n}}.$$

Тогда $\|f(x) - f(y)\|_{\mathbb{R}^n} \leqslant S\,\|x - y\|_{\mathbb{R}^m}$.

Обсудить теоремы о среднем. Сказать, что справедлива в любых банаховых пространствах (полные линейные нормированный пространства).

Принцип Банаха[3] сжимающих отображений.

Неподвижные точки, разрешимость уравнений. Примеры: непрерывное отображение отрезка в себя, непрерывное отображение шара или односвязной области в $\mathbb{R}^n$ в себя, отображение монотонным отображением отрезка в себя.

Пусть дано полное метрическое пространство X с метрикой $d(x,y)$.

Определение. Отображение $f : X \to X$ называется *сжимающим*, если оно удовлетворяет условию Липшица: $\forall x_1, x_2 \in X$ справедливо $d(f(x_1), f(x_2)) \leqslant q d(x_1, x_2)$ с константой Липшица $q < 1$.

Теорема. *Каждое сжимающее отображение f имеет единственную неподвижную точку $x_* = f(x_*)$.*

Доказательство. Рассмотрим произвольную точку $x_0 \in X$ и построим последовательность $x_{n+1} = f(x_n)$, $n \geqslant 0$. Эта последовательность фундаментальна.

Для доказательства фундаментальности напишем цепочку неравенств:

$$d(x_\ell, x_{\ell+1}) \leqslant q d(x_{\ell-1}, x_\ell) \leqslant q^2 d(x_{\ell-2}, x_{\ell-1}) \leqslant \ldots \leqslant q^\ell d(x_0, x_1).$$

Теперь для всех n, k

$$d(x_n, x_{n+k}) \leqslant \sum_{\ell=n}^{n+k-1} d(x_\ell, x_{\ell+1}) \leqslant \sum_{\ell=n}^{n+k-1} q^\ell d(x_0, x_1) = q^n(1 + q + q^2 + \ldots) d(x_0, x_1) \leqslant d(x_0, x_1)\frac{q^n}{1-q}.$$

Отсюда следует фундаментальность последовательности x_n, и, в силу полноты X, её сходимость к некоторому x_*. Теперь в силу непрерывности f имеем $x_* = \lim x_n = \lim x_{n+1} = \lim f(x_n) = f(\lim x_n) = f(x_*)$.

Нет двух предельных точек, очевидно: $d(x^*, x_*) = d(f(x^*), f(x_*)) \leqslant q\, d(x^*, x_*) \ \Rightarrow x^* = x_*$. $\square$

[3] Банах, Стефан, 1892-1945, замечательный польский математик. Придумал несколько вечных концепций (например, банахово пространство) и методов (например, метод сжимающих отображений)

Мы доказали в принципе сжимающих отображений, что последовательность $x_n = f(x_{n-1})$ при любом начальном значении x_0 сходится к единственной неподвижной точке x^*. В доказанной оценке

$$d(x_n, x_{n+k}) \leqslant d(x_0, x_1)\frac{q^n}{1-q}$$

можно перейти к пределу при $k \to \infty$:

$$d(x_n, x^*) \leqslant d(x_0, x_1)\frac{q^n}{1-q} \qquad \Big(\text{при } \boldsymbol{n = 0: \quad d(x_0, x^*) \leqslant d(x_0, x_1)\frac{1}{1-q}}\Big).$$

Эта оценка означает, что сходимость последовательных приближений в условиях теоремы о сжимающих отображениях сходится экспоненциально, как q^n.

Обобщение. *Пусть некоторая степень f^n непрерывного отображения f — сжимающее отображение. Тогда f имеет единственную неподвижную точку $x_* = f(x_*)$.*

Для доказательства сначала рассмотрим неподвижную точку $x = f^n(x) = \lim_{k\to\infty} f^{kn}(x_0)$ отображения f^n. Покажем, что $x = f(x)$.

В силу непрерывности f имеем $f(x) = \lim_{k\to\infty} f^{kn}(f(x_0))$. Теперь

$$d\big(f^{kn+1}(x_0), f^{kn}(x_0)\big) \leqslant q\, d\big(f^{(k-1)n+1}(x_0), f^{(k-1)n}(x_0)\big) \leqslant \ldots \leqslant q^k d(f(x_0), x_0).$$

Следовательно, $d(f^{kn+1}(x_0), f^{kn}(x_0)) \to 0 \;\Rightarrow\; f(x) = x$. □

Теорема об устойчивости неподвижной точки. Пусть отображение $f(x) = f(x, t) : X \to X$ зависит от параметра $t \in D$, где D — метрическое пространство с метрикой ρ (не компактное и даже не обязательно полное). Пусть снова при всех значениях t отображения $f(x) = f(x, t) : X \to X$ сжимающие с общей константой Липшица $q < 1$.

Пусть при каждом x отображение $f(x, t)$ как функция от t непрерывна в точке t_0.

Обозначим через $x^*(t)$ неподвижную точку отображения f.

Теорема. *Функция $t \mapsto x^*(t)$ непрерывна в точке t_0.*

Доказательство. При каждом t решение уравнения $x = f(x, t)$ может быть получено как предел последовательности $x_{n+1} = f(x_n, t)$, причем начальная точка x_0 может выбираться какая угодно. Пусть $x_0 = x^*(t_0) = f(x^*(t_0), t_0)$. Так как

$$d(x^*(t), x^*(t_0)) = d(x^*(t), x_0) \leqslant \frac{1}{1-q} d(x_1, x_0) = \frac{1}{1-q} d(f(x^*(t_0), t), f(x^*(t_0), t_0)) \to 0.$$

Здесь $x_1 = f(x_0, t)$, неравенство следует из доказанной оценки $d(x_n, x_{n+k}) \leqslant d(x_0, x_1)\dfrac{q^n}{1-q}$ при $n = 0$. □

Аналогия с линейными уравнениями. Пусть в полном нормированном пространстве, например, в $\mathbb{R}^n$, дано линейное уравнение $x = Ax + b$. Тогда единственное решение существует, если и только если 1 не является собственным значением линейного оператора A, оно определяется соотношением $x = (Id - A)^{-1}b$.

Пусть $\|A\| < 1$. Тогда для отображения $(Id - A)^{-1}$ справедлива формула

$$(Id - A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

Для доказательства достаточно умножить обе части равенства на невырожденную матрицу $Id - A$. Формула является полным аналогом суммы бесконечной геометрической прогрессии. Соответственно, для решения уравнения верна формула

$$(Id - A)^{-1}b = \sum_{k=0}^{\infty} A^k b.$$

Если взять любое начальное приближение x_0, а потом положить $x_n = Ax_{n-1} + b$, то получатся те самые приближения, что и в нелинейном случае: $x_0,\ Ax_0 + b,\ A^2x_0 + Ab + b,\ A^3x_0 + A^2b + Ab + b$ и так далее. Так как $A^k x_0 \to 0$, то получается аналогия между линейным и нелинейным случаем.

Лекция 9 (16 мая 2016)

На прошлой лекции мы рассмотрели несколько важных вещей.

1) Теорема о конечных приращениях.

2) Норма линейного оператора.

3) Принцип сжимающих отображений, в том числе параметрические варианты.

Это всё были, как мне кажется, очень важные вещи: и условие Липшица для отображения с ограниченной производной, и принцип Банаха сжимающих отображений.

Теорема о неявной функции. Рассмотрим пространства $X = \mathbb{R}^m$ и $Y = \mathbb{R}^n$ и уравнение $f(x, y) = 0$, где отображение f со значениями в Y определено и непрерывно на произведении каких-то подмножеств $D \subset X$ и $G \subset Y$.

Пусть (x_0, y_0) — внутренняя точка области определения f и $f(x_0, y_0) = 0$, пусть отображение f дифференцируемо по y при каждом x, пусть «частная» производная $f'_y(x_0, y_0) : Y \to Y$ обратима. Пусть ещё производная f'_y непрерывно в точке (x_0, y_0) зависит от x и y, как функция из $D \times G \to \mathcal{L}(Y, Y)$.

Последнее условие означает непрерывность функции $(x, y) \mapsto \mathcal{L}(X, Y)$. Конечно, норма в конечномерном пространстве $\mathcal{L}(X, Y)$ при $X = \mathbb{R}^m$ и $Y = \mathbb{R}^n$ может быть выбрана любой, они все эквивалентны, однако про себя мы помним, что норма в этом пространстве операторная, как вы знаете. Это же можно перефразировать так, что все частные производные матрицы Якоби $\partial f_i / \partial y_j$ — непрерывны. Это означает непрерывность покоординатную, она эквивалентна всем остальным.

Теорема. *Существует непрерывная неявная функция.*

1) Уравнение $f(x, y) = 0$ в Y эквивалентно уравнению $\big(f'_y(x_0, y_0)\big)^{-1} f(x, y) = 0$. Здесь обратимый линейный оператор $\big(f'_y(x_0, y_0)\big)^{-1}$ (он не зависит ни от x, ни от y) применяется к f, его значение 0 только в начале координат.

2) Уравнение $\big(f'_y(x_0, y_0)\big)^{-1} f(x, y) = 0$ эквивалентно уравнению $y = T(x, y)$, где $T(x, y) = y - \big(f'_y(x_0, y_0)\big)^{-1} f(x, y)$. Очевидно.

3) Производная T'_y в точке (x_0, y_0) равна 0 — это легко посчитать:

$$T'_y(x_0, y_0) = Id - \big(f'_y(x_0, y_0)\big)^{-1} f'_y(x_0, y_0) = 0.$$

4) Так как производная f'_y непрерывна в точке (x_0, y_0), то можно выбрать такие окрестности $\|x - x_0\|_X \leqslant \rho$ по x и $\|y - y_0\|_Y \leqslant r$ по y точки (x_0, y_0), что $\|T'_y(x, y)\|_{\mathcal{L}(Y,Y)} \leqslant 1/4$.

5) По теореме о среднем $\|T(x, y_1) - T(x, y_2)\|_Y \leqslant (1/4)\|y_1 - y_2\|_Y$, то есть оператор T сжимающий. Значит, при каждом x существует единственный y, по доказанному ранее этот y непрерывно зависит от x.

6) Пока что мы не увидели, что какое-то множество (метрическое пространство) отображается в себя.

7) Мы будем выбирать в качестве метрического пространства шар в Y с центром в y_0 и радиуса r. Так как $y_0 = T(x_0, y_0)$ и

$$\|y_0 - T(x,y)\|_Y = \|T(x_0,y_0) - T(x,y)\|_Y \leqslant \|T(x_0,y_0) - T(x,y_0)\|_Y + \|T(x,y_0) - T(x,y)\|_Y,$$

то при $\|T(x_0,y_0) - T(x,y_0)\|_Y \leqslant r/4$ (а так будет при достаточно малом $\|x - x_0\|_X$)

$$\|y_0 - T(x,y)\|_Y \leqslant r/4 + (1/4)\|y - y_0\|_Y \leqslant r$$

и шар $\|y - y_0\|_Y \leqslant r$ преобразуется оператором T в себя.

8) Так как $T(x,y)$ зависит от x непрерывно, а константа Липшица $q = 1/4$ общая, то по параметрическому варианту метода сжимающих отображений функция $y(x)$ — непрерывная. $\square$

Замечания.

1. Можно было бы такую теорему о неявной функции доказать, рассмотрев оператор в пространстве непрерывных функций.

2. Для доказательства разрешимости уравнения используется метод неподвижной точки. Есть еще всякие разные методы, я сейчас скажу об одном из них, в Зориче все написано подробно, 1 страница.

Дано уравнение $F(x,y) = 0$. Рассмотрим скалярнозначную функцию $\varphi(x,y) = \|F(x,y\|_Y^2$, сумму квадратов компонент. В точке (x_0, y_0) эта функция обращается в ноль, а окрестности точки (x_0, y_0) (так как F'_y не вырожденная матрица), эта функция примерно квадратичная.

Поэтому при каждом x из окрестности у нее есть единственный минимум, в котором значение квадратичной функции равно нулю. Вот так и определяется неявная функция.

Такой подход называется вариационным подходом и используется в разных областях современной математики. В частности, он позволяет доказывать разрешимость всяких хитрых дифференциальных уравнений и краевых задач.

3. Замечание про линейный случай. Пусть отображение F — линейное, $x_0 = y_0 = 0$.

Это значит, что задана система из n уравнений с $n+m$ неизвестными, m штук называются x и n штук называются y. Утверждение теоремы о неявной функции означает, что если определитель по y отличен от нуля, то y можно выразить через x. В линейном случае это факт, который вы проходили по линейной алгебре: есть уравнение $Ax + By = 0$, матрица B — квадратная невырожденная, тогда определено отображение $y = -B^{-1}Ax$, при каждом x соответствующий y решает уравнение.

4. Рассуждение про формулу для производной. Я доказал только что теорему о существовании непрерывной неявной функции, при этом я использовал только существование непрерывной обратимой производной по y. Пусть $F \in C^1$ непрерывно дифференцируемо по x и по y.

Тогда из уравнения $F(x,y)=0$ следует равенство $F'_x(x,y)(x-x_0)+F'_y(x,y)(y-y_0)+$ малые слагаемые $=0$. Таким образом, $y=y_0-\big(F'_y(x,y)\big)^{-1}F'_x(x,y)(x-x_0)+$ малые слагаемые. Это и есть формула для производной y'_x.

$$\begin{aligned}
0 &= f(x+h,y(x+h))-f(x,y(x)) = \\
&= f(x+h,y(x+h))-f(x,y(x+h))+f(x,y(x+h))-f(x,y(x)) = \\
&= f'_x(x,y(x+h))h+o(h)+f'_y(x,y(x))(y(x+h)-y(x))+o(y(x+h)-y(x)) = \\
&= f'_x(x,y)h+f'_y(x,y)(y(x+h)-y(x))+o(h)+o(y(x+h)-y(x)).
\end{aligned}$$

Поэтому,

$$y(x+h)-y(x)=(f'_y(x,y))^{-1}f'_x(x,y)h+o(h)+o(y(x+h)-y(x)).$$

Теперь отсюда, во-первых, следует, что $y(x+h)-y(x)=O(h)$, и, во-вторых, требуемая формула.

5. Теперь я сформулирую теорему о неявной функции ещё раз, ту же что я формулировал пару дней назад.

Теорема о неявной функции. *Пусть задано отображение $F:U\subset\mathbb{R}^{m+n}\to\mathbb{R}^n$, U — окрестность точки (x_0,y_0). Пусть*

1) $F(x_0,y_0)=0$;

2) $F\in C^p(U,\mathbb{R}^n),\ p\geqslant 1$;

3) Дифференциал $F'_y(x_0,y_0)$ — обратимый линейный оператор;

Тогда существуют промежутки $I^m_x=\{x\in\mathbb{R}^m:\|x-x_0\|\leqslant\alpha\}$ и $I^n_y=\{y\in\mathbb{R}^n:\|y-y_0\|\leqslant\beta\}$ и такое отображение $f\in C^p(I^m_x,I^n_y)$, что что для любой точки $(x,y)\in I^m_x\times I^n_y$ справедливо равенство

$$F(x,y)=0\quad\Leftrightarrow y=f(x),\qquad f'(x)=-\big(F'_y(x,f(x))\big)^{-1}F'_x(x,f(x)).$$

6. Существование старших производных следует из доказанной формулы про первую производную.

Теперь докажем ранее сформулированное утверждение о дифференцируемости функции у которой есть непрерывные частные производные.

Утверждение. *Если функция дифференцируемая в точке x, то существуют частные производные в точке x. Если функция имеет частные производные в окрестности точки x, и они непрерывны в x, то она дифференцируема в x.*

Теорема. *Если функция имеет частные производные в окрестности точки x, и они непрерывны в x, то она дифференцируема в x.*

Докажу это для функции двух переменных, для общего случая чуть более громоздко, но точно так же. Дана функция f, непрерывная в окрестности U точки $x=(x_0,y_0)$. Пусть в окрестности U существуют частные производные f'_x и f'_y, причём они непрерывны в точке (x_0, y_0). Пишем

$$\begin{aligned}\Delta f &= f(x_0+\Delta_x, y_0+\Delta_y) - f(x_0, y_0) = \\ &= f(x_0+\Delta_x, y_0+\Delta_y) - f(x_0+\Delta_x, y_0) + f(x_0+\Delta_x, y_0) - f(x_0, y_0) = \\ &= f'_y(x_0+\Delta_x, y_0+\theta_y\Delta_y)\Delta_y + f'_x(x_0+\theta_x\Delta_x, y_0)\Delta_x\end{aligned}$$

Теперь $\Delta f - f'_x(x_0, y_0)\Delta_x - f'_y(x_0, y_0)\Delta_y = \Delta_x o(1) + \Delta_y 0(1)$. □

Старшие производные.

Что такое вторая производная отображения? Это должна быть производная от производной.

То есть конструкция должна быть такая. Вот есть отображение f, где-то определенное на $\mathbb{R}^m$ и со значениями в каком-то пространстве $\mathbb{R}^n$. Пусть оно в каждой точке дифференцируемое. Тогда в каждой точке определена производная $f'(x)$, это линейный оператор на соответствующих касательных пространствах.

Соответственно, определено отображение $x \mapsto f'(x)$ из пространства $\mathbb{R}^m$ в пространство линейных операторов $\mathcal{L}(\mathbb{R}^m_x, \mathbb{R}^n_x)$. Это конечномерное линейное пространство, его размерность равна mn.

Если там ввести удобную метрику, то можно исследовать непрерывность оператора $x \mapsto f'(x)$, его свойства, можно изучать его дифференцируемость.

Соответственно, вторая производная отображения $x \mapsto f'(x)$ — это будет линейный оператор из $\mathbb{R}^m$ в $\mathbb{R}^{mn}$... и так далее. Мы пока не будем изучать эти конструкции, они в этом модуле нам почти не понадобятся.

Вторую производную $f''(x_0)$ естественно представлять как билинейное отображение $f''h_1h_2$, применяем $f''(x_0)$ к вектору из касательного пространства, получаем линейный оператор. применяем его к другому вектору из касательного пространства — получаем вектор из $\mathbb{R}^n$ (из соответствующего другого касательного пространства).

Понадобится только пространство $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ и норма в нём.

Однако, если предположить, что во всех пространствах есть система координат, то можно рассматривать обычные частные производные любых порядков и в любом сочетании.

Я буду обозначать их по-разному, частная производная функции $f(x_1, \ldots, x_m)$ по переменной x_k будет обозначаться

$$f'_{x_k}, \quad \partial_k f, \quad \frac{\partial f}{\partial x_k} \quad \text{и даже (чтобы штрихи не писать!)} \quad f_{x_k}.$$

Их легко определить для $f : \mathbb{R}^m \to \mathbb{R}$ по индукции: если определена производная $\partial_{i_1,\ldots,i_k} f$ функции f порядка k по переменным $i_1, \ldots, i_k$ (индексы могут повторяться), то производную $(k+1)$-го порядка по переменной x_s определим равенством $\partial_{s,i_1,\ldots,i_k} f = \partial_s(\partial_{i_1,\ldots,i_k} f)$.

Старшие производные перестановочны.

Теорема. *Пусть функция $f : \mathbb{R}^m \to \mathbb{R}$ имеет в области $G \subset \mathbb{R}^m$ частные производные второго порядка $\partial_{i_1,i_2} f$ и $\partial_{i_2,i_1} f$. Если обе эти производные непрерывны в некоторой точке $x \in G$, то они совпадают в этой точке.*

Доказательство этой теоремы не слишком громоздкое, оно опирается на формулу Лагранжа. Проведу его для функции двух переменных.

Есть функция f непрерывная в окрестности точки (x, y), существуют производные f''_{xy} и f''_{yx}, причём они непрерывны в точке (x, y). Рассмотрим вспомогательную функцию

$$F(h_1, h_2) = f(x + h_1, y + h_2) - f(x + h_1, y) - f(x, y + h_2) + f(x, y).$$

Положим $\varphi(t) = f(x + th_1, y + h_2) - f(x + th_1, y)$, тогда $F(h_1, h_2) = \varphi(1) - \varphi(0)$ и

$$F(h_1, h_2) = \varphi'(\theta) = f'_x(x + \theta h_1, y + h_2) - f'_x(x + \theta h_1, y)h_1 = f''_{yx}(x + \theta h_1, y + \eta h_2)h_1 h_2.$$

Аналогично, положив $\psi(t) = f(x + h_1, y + th_2) - f(x, y + th_2)$, получим $F(h_1, h_2) = \psi(1) - \psi(0)$ и, снова, $F(h_1, h_2) = f''_{xy}(x + \bar{\theta}h_1, y + \bar{\eta}h_2)h_1 h_2$. Отсюда и из непрерывности следует утверждение теоремы. $\square$

Замечу, что условия теоремы все существенные. Нельзя отказаться от непрерывности, нельзя отказаться от дифференцируемости в окрестности.

В частности, обращаю внимание, что в теореме используются не только переменные с индексами i_1 и i_2. Неявно используются и все остальные переменные: все функции и производные обязаны быть непрерывными по всем переменным.

Следствие. *Если функция $f \in C^k(G; \mathbb{R})$, то значение частной производной $\partial_{i_1,\dots,i_k} f(x)$ не зависит от порядка индексов.*

Лекция 10 (23 мая 2016)

На прошлой лекции была теорема о неявной функции, теорема о перестановках порядка дифференцирования в смешанных производных.

Теперь мы сделаем некоторые выкладки, которые позволят нам сформулировать формулу Тейлора для скалярнозначных функций многих переменных.

1. Пусть дана функция $f(x_1, x_2)$ класса C^k в некоторой области G (это значит, что у функции есть все частные производные до порядка k включительно, причем все они непрерывны).

Пусть $h = (h_1, h_2)$ такое, что отрезок $[x, x+h]$ содержится области G. рассмотрим функцию

$$\varphi(t) = f(x + th) : [0, 1] \to \mathbb{R}, \quad f \in C^k.$$

Посчитаем производные этой функции с помощью теорем о производной сложной функции:

$$\varphi'(t) = \partial_1 f(x_1 + th_1, x_2 + th_2)h_1 + \partial_2 f(x_1 + th_1, x_2 + th_2)h_2$$

(здесь обозначено $\partial_i = \partial/\partial x_i$)

$$\varphi''(t) = \partial_{11} f(x + th)(h_1)^2 + 2\partial_{12} f(x + th)h_1 h_2 + \partial_{22} f(x + th)(h_2)^2.$$

Если аккуратно еще раз продифференцировать, получится

$$\varphi^{(p)}(t) = (h_1\partial_1 + h_2\partial_2)^p f(x + th) = (\nabla, h)^p f(x + th).$$

2. Выражение $(\nabla, h)^p f$ — это форма записи. Здесь дифференциальные операторы (в этом термин я не вношу никакого глубокого смысла, просто — оператор, содержащий дифференцирование) $(\nabla, h)^p$ осмысленные, так как операторы ∂_ℓ коммутируют для разных ℓ.

3. Теперь пусть функция от m переменных. Тогда снова $\varphi(t) = f(x+th) : [0, 1] \to \mathbb{R}, \quad f \in C^k$, и формулу также можно переписать в виде

$$\varphi^{(p)}(t) = \Big(\sum_{i=1}^{m} h_i\partial_i\Big)^p f(x + th) = (\nabla, h)^p f(x + th).$$

4. Есть еще один способ записывать такие громоздкие формулы. Считается, что если индекс входит в формулу 2 раза, причем один раз в виде нижнего индекса, а другой раз — в виде верхнего индекса, то по этому индексу происходит суммирование. Это часто бывает удобно, если верхний индекс нельзя спутать с показателем степени и если из контекста ясен диапазон суммирования. Формулу для производных функции φ можно переписать тогда

$$\varphi^{(p)}(t) = h^{i_1} \dots h^{i_p} \partial_{i_1,\dots,i_p} f(x + th).$$

Здесь индексы у координат вектора h подняты наверх, суммирование идет по p переменным i_p.

Теперь перейдем к формуле Тейлора.

Теорема. *Пусть* $f : U(x) \to \mathbb{R}$, $U \subset \mathbb{R}^m$ *и пусть* $f \in C^k$, *и пусть* $[x, x+h] \in U$. *Тогда*

$$f(x+h) - f(x) = \sum_{\ell=k-1}^{\ell-1} \frac{1}{\ell!}(\nabla, h)^\ell f(x) + r_{k-1}(x,h), \qquad r_{k-1}(x,h) = \int_0^1 \frac{(1-t)^k}{(k-1)!}(\nabla, h)^k f(x+th)\, dt.$$

Доказательство. Рассматриваем функцию $\varphi(t) = f(x+th) \in C^k$, для неё пишем формулу Тейлора (это функция одной переменной)

$$\varphi(t) = \varphi(0) + \frac{\varphi'(0)}{1!}t + \ldots + \frac{\varphi^{(k-1)}(0)}{(k-1)!}t^{k-1} + \int_0^t \frac{(t-\tau)^k}{(k-1)!}\varphi^{(k)}(\tau)\, d\tau.$$

Я рассказывал эту формулу сразу после интегрирования по частям. Теперь полагаем $t = 1$ и пользуемся формулами для производных функции φ, которые я раньше предусмотрительно получил. $\square$

Естественно, из интегральной формулы для остаточного члена $r_{k-1}(x,h)$ вытекают иные формы. Например, форма Лагранжа:

$$r_{k-1}(x,h) = \frac{1}{k!}(\nabla, h)^k f(x+\theta h)$$

(используем теорему о среднем) или форма Пеано: $r_{k-1}(x,h) = o(\|h\|^k_{\mathbb{R}^m})$.

Экстремумы функций многих переменных.

Пусть задана функция $f : U(x) \to \mathbb{R}$, $U \subset \mathbb{R}^m$. Мы будем думать про себя, что эта функция гладкая, для определений это не нужно, но мы будем считать, что у нее есть «нужное количество» непрерывных производных. Иногда одна, иногда две.

Теперь мы займемся изучением минимумов и максимумов этой функции, а также седловых точек функций многих переменных.

Определения. 1. *Локальный максимум.* Если у точки x_0 есть окрестность, в которой выполнено неравенство $f(x) \leqslant f(x_0)$, то это локальный максимум. Если в проколотой окрестности выполнено более сильное неравенство $f(x) < f(x_0)$, то это строгий локальный максимум.

2. *Локальный минимум* и строгий локальный минимум. Определяется совершенно аналогично.

3. *Локальный экстремум.* Общее слово для минимумов и максимумов — экстремум.

4. *Критическая точка функции.* Точка x называется критической точкой функции, если $\mathbf{grad}\, f(x) = 0$, то есть $f'_{x_i}(x) = 0$ при всех $i = 1, \ldots, m$.

Критические точки также называют *стационарными* точками.

5. *Седловая точка.* Обезьянье седло. Критическая точка функции двух переменных, не являющаяся экстремумом, является-называется седловой точкой. Седловая точка является аналогом точки перегиба для функций одной переменной.

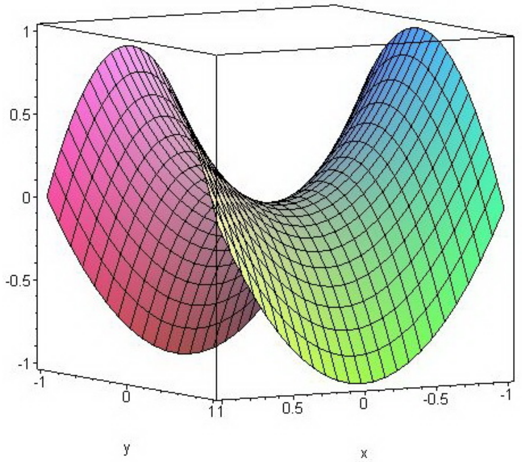


Рис. 2

Это график функции $z = x^2 - y^2$. Начало координат — седловая точка.

6. Глобальный минимум. **Вопрос к залу:** может ли быть гладкая функция на плоскости, у которой есть ровно одна критическая точка, это точка локального минимума, и чтобы этот локальный минимум не был глобальным?

В случае скалярных функций так не бывает, а для плоскости — бывает.

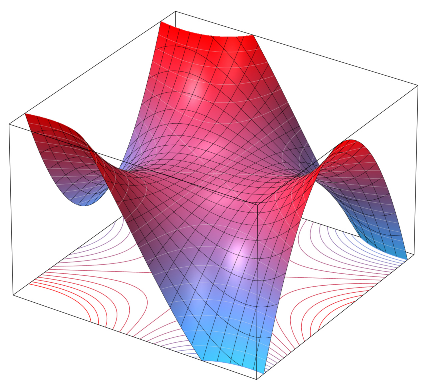


Рис. 3

Это график функции $z = x^3 - 3xy^2$. Начало координат — обезьянье седло.

Теорема (необходимое условие локального экстремума). *Пусть функция* $f : \mathbb{R}^m \to \mathbb{R}$ *имеет во внутренней точке* x *области определения все частные производные (по каждой из переменных). Пусть у* f *точка* x *является локальным экстремумом. Тогда* $\partial_i f(x) = 0,\ i = 1, \dots, m$.

Выберем $i = 1, \dots, m$ и рассмотрим скалярную функцию $\psi(t) = f(x_1, \dots, x_i + t, \dots, x_m)$. Эта функция по предположению имеет при $t = 0$ локальный экстремум, ее производная $\psi'(0) = 0$, поэтому $\partial_i f(x) = 0$. $\square$

Иными словами, каждая точка экстремума дифференцируемой функции — критическая. Обратное, конечно, не верно. У скалярной функции $f(x) = x^3$ ноль — критическая точка, не являющаяся локальным экстремумом. Седловые точки не являются точками локального экстремума.

Из этой теоремы естественным путем следует возможность нахождения локальных экстремумов: пишем систему из m скалярных уравнений $\mathbf{grad}\, f(x) = 0$ с m неизвестными (координатами точки x). Находим решения и дальше как-то надо исследовать полученные точки.

В скалярном случае, если помните, удавалось воспользоваться второй производной: если $f'' > 0$, то это был минимум, если $f'' < 0$, то это был максимум, случай $f'' = 0$ требовал рассмотрения старших производных.

В следующей теореме надо вспомнить, что такое положительно определённые симметрические матрицы и что такое отрицательно определённые симметрические матрицы:

$$(Ax, x) > 0 \ \ (x \neq 0) \text{ — это положительно определённая матрица,}$$

$$(Ax, x) < 0 \ \ (x \neq 0) \text{ — это отрицательно определённая матрица.}$$

Причем не $\geqslant$, $\leqslant$, а строгие неравенства.

Это означает, что собственные числа матрицы (а у симметрических матриц они всегда вещественные) все положительны или все отрицательны.

Теорема (достаточное условие локального экстремума). *Пусть $f \in C^2$, пусть точка x_0 — критическая. Рассмотрим квадратичную форму*

$$A(h) = (f''(x_0)h, h) = \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h^i h^j \equiv \partial_{ij} f(x_0) h^i h^j$$

переменных h^ℓ. Если эта квадратичная форма определена положительно, то точка x_0 является строгим локальным минимумом. Если эта квадратичная форма определена отрицательно, то точка x_0 является строгим локальным максимумом. Если форма принимает значения разных знаков, то точка x_0 не является точкой экстремума.

Квадратичная форма определена положительно, это значит, что её значения всюду положительны, кроме нуля. Форма x^2 на плоскости (x, y) не является положительной: она равна нулю при $x = 0$ и любом y. Такие случаи теорема не охватывает.

В случае таких «не строго» определённых квадратичных форм сделать вывод (экстремум или нет) нельзя. Примеры легко придумать: $f(x, y) = x^2 - y^4$ и $f(x, y) = x^2 + y^4$, в обоих этих примерах форма $A(h_1, h_2) = h_1^2$ не является строго положительной, в одном случае в нуле локальный минимум, в другом — седло.

Доказательство. Рассмотрим форму A на сфере $h \in S(0,1) \in \mathbb{R}^m$. Так как A непрерывная функция и так как $S(0,1)$ — компактное множество, то существуют конечные величины $m = \inf A(h)$ и $M = \sup A(h)$.

Запишем формулу Тейлора с остаточным членом в форме Пеано:

$$f(x_0 + h) - f(x_0) = \frac{\|h^2\|}{2!}\big(A(h) + o(1)\big).$$

Путь квадратичная форма A положительно определена. Это значит, что $m > 0$, следовательно при малых $\|h\| > 0$ правая часть формулы Тейлора положительна, значит, точка x_0 является строгим локальным минимумом.

Аналогичное рассуждение в случае, когда квадратичная форма A отрицательно определена. В этом случае $M < 0$, и точка x_0 является строгим локальным максимумом.

Если форма принимает и положительные, и отрицательные значения, то $m < 0 < M$. В этом случае существуют точки $h_-, h_+ \in S(0,1)$, в которых $\pm A(h_\pm) > 0$. Тогда при достаточно малых $\theta > 0$ при $h = \theta h_\pm$ знак правой части формулы Тейлора совпадает со знаком формы $A(h_\pm)$ то есть принимает и отрицательные, и положительные значения. Поэтому x_0 не является экстремумом. □

Как определить, является ли форма положительно определённой? Есть критерий Сильвестра (форма положительно определена, если и только если все главные миноры положительны), можно найти все собственные значения и увидеть, что они все положительные, но это все вычислительно сложные задачи.

Метод наименьших квадратов. Линейная регрессия. Я быстро расскажу об одном приложении полученных теорем. Когда-то меня этому научили врачи из кардиологического центра, конечно, сразу оказалось что это все знают. Но я после мехмата как-то не знал.

Пусть вы врач, у вас есть данные каких-то измерений, вы меряете 2 величины x и y (или 3, или m), но я буду считать, что их 2. И делаете n одновременных измерений этих величин.

По вашим медицинским соображением вы считаете, что эти величины связаны между собой линейно: $y = ax + b$. Задача: найти коэффициенты a и b, связывающие ваши величины.

Обычно, в «практической жизни» это делают так. Врач нажимает кнопки на компьютере, вводит свои данные, 2 вектора x_i и y_i, $i = 1, \ldots, n$. А компьютер выдают врачу приблизительные значения коэффициентов a и b, которые связывают величины x и y. Компьютер выдаёт ещё несколько величин, описывающих надежность определения этих коэффициентов, но это выходит за рамки нашего повествования.

Компьютер считает оптимальные значения a и b таким методом. Составляем сумму

$$f(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2.$$

Это квадратичная форма, у нее есть минимум (мы это знаем по смыслу), для его определения надо написать 2 уравнения $f'_a = f_b = 0$ с неизвестными a и b, эти уравнения имеют вид

$$\sum_{i=1}^{n} 2(y_i - ax_i - b)x_i = \sum_{i=1}^{n} 2(y_i - ax_i - b) = 0.$$

Это система линейных уравнений $a\sum x_i^2 + b\sum x_i = \sum x_i y_i, \qquad a\sum x_i + nb = \sum y_i$. Решаем, находим значения a и b.

Диффеоморфизмы.

Теперь мы временно откладываем экстремумы и начинаем заниматься приложениями теоремы о неявной функции. Это будет ряд довольно громоздких, хотя и философски очень простых утверждений. Начнём с мегаважного определения.

Определение. *Отображение $f : U \to V$, областей $U, V \subset \mathbb{R}^m$ называется* **диффеоморфизмом класса** p, *если выполнены 3 условия:*

1) f *— взаимно однозначное отображение (биекция);*

2) $f \in C^p(U, V)$,

3) $f^{-1} \in C^p(V, U)$.

Диффеоморфизм класса 0 называется ***гомеоморфизм***.

Примеры.

1. Линейное отображение является диффеоморфизмом, если и только если матрица квадратная и определитель отличен от нуля, вроде это очевидно.

2. Отображение $x \mapsto x^3$ любой области в $\mathbb{R}$, содержащей 0, является гомеоморфизмом, но не является диффеоморфизмом класса 1.

3. Проверка условий 1 и 3 сложная!

Теорема о диффеоморфизме (об обратной функции). *Если $f \in C^p(\mathbb{R}^m, \mathbb{R}^m)$, $p > 0$ — отображение, причем $y_0 = f(x_0)$ и $f'(x_0)$ — обратимое линейное отображение, то существуют окрестности точек x_0 и y_0, в окрестности которых f является диффеоморфизмом класса p, при этом $\left(f^{-1}\right)'(y) = \left(f'\right)^{-1}(x)$.*

Доказательство простое — это следствие теоремы о неявной функции.

Рассмотрим уравнение $y = f(x) \;\Leftrightarrow\; F(x,y) = f(x) - y = 0$. Мы хотим разрешить это уравнение относительно переменной x в окрестности некоторой точки $(x_0, y_0) \in \mathbb{R}^m \times \mathbb{R}^m$.

По предположению выполнены все условия теоремы о неявной функции, $F'_x(x_0, y_0) = f'(x_0)$ — обратимое отображение.

В силу теоремы о неявной функции отсюда следует существование отображения $x = g(y)$, удовлетворяющего $x = g(y) \;\Leftrightarrow\; y = f(x)$, отсюда же следует требуемая гладкость $g = f^{-1}$ и равенство про производные:

$$g'(y) = -\big(F'_x(x,y)\big)^{-1}\big(F'_y(x,y)\big),\; F'_x(x,y) = f'(x),\; F'_y(x,y) = -Id \quad\Rightarrow\quad g'(y) = \big(f'(x)\big)^{-1}.$$

Окрестности выбираются в соответствии с теоремой о неявной функции. $\square$

Эту теорему следует понимать как критерий локального диффеоморфизма: если есть гладкое отображение и его производная в некоторой точке обратима, то это диффеоморфизм.

Естественно, при этом должны совпадать размерности области определения и пространства образа, иначе не возможны условия теоремы (обратимы только квадратные матрицы).

Одномерный диффеоморфизм класса C^1 — это всегда гладкая функция с ненулевой производной. Это следствие обратимости и дифференцируемости и функции, и обратной функции.

Правильно понимать также эту теорему как аналог соответствующего линейного утверждения: если линейное отображение обратимо, то оно — диффеоморфизм. Локально гладкое отображение ведет себя примерно как его дифференциал.

Лекция 11 (25 мая 2016)

На прошлой лекции была формула Тейлора и условия локального экстремума, необходимые и достаточные. Необходимые — это должна быть критическая точка, а достаточные — это в критической точке должна была быть знакоопределена квадратичная форма из вторых производных. Кроме того, мы начали новую тему про диффеоморфизмы. Дали определение диффеоморфизма и сформулировали теорему о диффеоморфизме, она же теорема об обратной функции.

Диффеоморфизмы используются (в частности) для замены переменных.

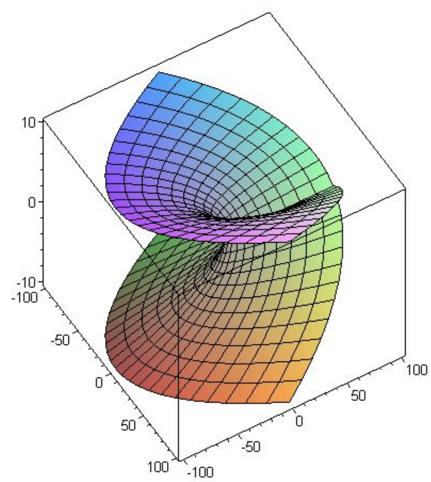
Если есть диффеоморфизм $\phi : U \to V$, то говорят, что на U заданы другие координаты. Был вектор координат u, а теперь можно говорить о координатах v, зная, что u вычисляются через v однозначно. Обычно все это делается локально, в окрестности некоторой точки.

Приведу пример использования теоремы о диффеоморфизме для выпрямления кривых.

Критическая точка отображения. Пусть задано отображение $f : \mathbb{R}^m \to \mathbb{R}^n$, причём пусть в обоих пространствах зафиксированы базисы. Тогда производная f' — это её матрица Якоби. Вообще говоря, естественно ожидать, что ранг матрицы размером $m \times n$ равен $\min\{m, n\}$. Точка называется критической, если ранг этой матрицы меньше $\min\{m, n\}$.

Критические точки функции (экстремумы, сёдла) были критическими в смысле этого определения: градиент равен нулю, то есть производная — матрица $1 \times m$ была нулевая, имела ранг равный 0.

Это понятие имеет чисто геометрический смысл.



Это образ квадрата $[-10, 10] \times [-10, 10] \subset \mathbb{R}^2$ при отображении $(u, v) \mapsto (uv, u^2 - v^2, v)$. Начало координат — критическая точка, в ней ранг матрицу Якоби равен 1, во всех остальных точках он равен 2.

В окрестности некритической точки образ имеет почти линейный–плоский вид, а в окрестности критической точки там может быть сложно устроенное ветвление.

Путь есть кривая на плоскости, заданная скалярным уравнением $F(x, y) = 0$. Пусть F — гладкая функция, пусть $F(x_0, y_0) = 0$ и пусть точка (x_0, y_0) не является критической: то есть ранг отображения $DF(x_0, y_0)$ равен 1, то есть по крайней мере одна из производных $F'_x(x_0, y_0)$ или $F'_y(x_0, y_0)$ отлична от нуля.

Я буду предполагать, что $F'_y(x_0, y_0) \neq 0$.

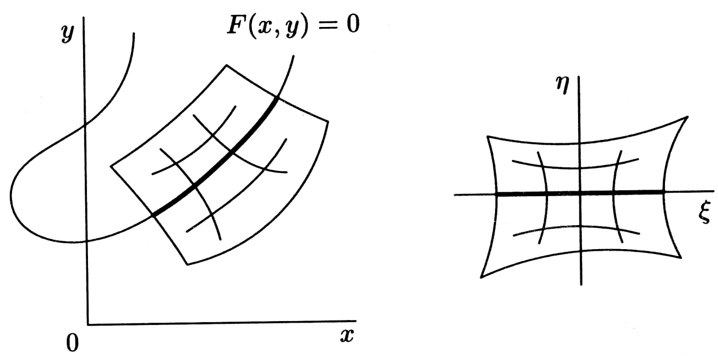
Я укажу такой диффеоморфизм окрестности точки (x_0, y_0) на окрестность нуля, что в новых координатах (ξ, η) кривая будет иметь уравнение существенно более простого вида: $\eta = 0$.

На вопрос: «Зачем это надо?» я не буду отвечать, однако, вроде ясно, что в каких-то конструкциях это удобно.

Положим $\xi = x - x_0,\ \eta = F(x, y)$. При таком отображении $\Phi : (x, y) \mapsto (\xi, \eta)$ точка (x_0, y_0) перейдет в начало координат, матрица Якоби этого отображения имеет вид

$$\begin{pmatrix} 1 & 0 \\ F'_x & F'_y \end{pmatrix},$$

якобиан равен $F'_y(x, y)$ и отличен от нуля в точке (x_0, y_0).



По доказанной теореме отображение Φ является диффеоморфизмом, у него есть обратное отображение $\Phi^{-1} : (\xi, \eta) \mapsto (x, y)$, кривая $F(x, y) = 0$ в новых координатах (ξ, η) имеет требуемый вид $\eta = 0$ (конечно, только локально, в некоторой окрестности начала координат).

Теперь пусть всё то же самое, уравнение от многих переменных в пространстве $\mathbb{R}$, то есть $F : \mathbb{R}^{m+1} \to \mathbb{R}$, одно скалярное уравнение, переменных $m + 1$ — множество решений — поверхность коразмерности 1. Будем считать, что $x \in \mathbb{R}^m,\ y \in \mathbb{R}$, уравнение имеет вид $F(x, y) = 0$.

Пусть $F(x_0, y_0) = 0$ и $F'_y(x_0, y_0) \neq 0$. Тогда отображение $\xi = x - x_0,\ \eta = F(x, y)$ имеет матрицу Якоби

$$\begin{pmatrix} Id_m & 0 \\ F'_x & F'_y \end{pmatrix},$$

Здесь Id_m — тождественное преобразование в $\mathbb{R}^m$, F'_x — вектор-строка, 0 — вектор-столбец с нулевыми координатами. Это — треугольная матрица и её определитель равен F'_y. Снова по теореме о диффеоморфизме отображение $(x, y) \mapsto (\xi, \eta)$ — диффеоморфизм, В координатах (ξ, η) поверхность приобретает локальный вид $\eta = 0$.

Пытливый слушатель спросит, что будет, если $F(x, y) = 0$, $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, $F(x_0, y_0) = 0$ и матрица $F'_y(x_0, y_0)$ обратима. А то же самое будет. В пространстве размерности 5 задано 3 уравнения. Фиксируем точку двумерной поверхности, предполагаем, что последние 3 переменных выбраны так, что матрица Якоби по последним трем переменным обратима. Снова есть диффеоморфизм окрестности в окрестности начала координат, такой что поверхность имеет вид $(\xi_1, \xi_2, 0, 0, 0)$.

Конечно, такое уже нарисовать не получится. И представить себе такое геометрически я не умею. Однако, аналитически, это хорошо видно. 5 уравнений, 3 хороших переменных, относительно них систему уравнений можно разрешить, 2 параметра остались. Вот и получился такой вид: $(\xi_1, \xi_2, 0, 0, 0)$.

Локальное приведение гладкого отображения к каноническому виду.

Напомню, что ранг матрицы — это размер самого большого невырожденного минора. Если мы умножим матрицу ранга k слева или справа на невырожденную квадратную матрицу, то ранг матрицы не изменится. Это вы должны были проходить по линейной алгебре и я считаю это известным.

Тем самым корректно говорить о ранге $\operatorname{rang} A$ линейного отображения $A : \mathbb{R}^m \to \mathbb{R}^n$, даже если базисы в пространствах $\mathbb{R}^m$ и $\mathbb{R}^n$ не выбраны и матрица отображения не определена.

Пусть задано нелинейное гладкое отображение $f : \mathbb{R}^m \to \mathbb{R}^n$. Будем называть *рангом* $\operatorname{rang} f(x)$ *отображения* f *в точке* x ранг его дифференциала $f'(x)$.

Естественно, ранг нелинейного отображения может меняться в зависимости от точки x. В нормальной ситуации $\operatorname{rang} f(x)$ локально не может уменьшиться. Ну в самом деле, определитель матрицы — непрерывная функция элементов матрицы, следовательно, если какой-то минор дифференциала отличен от нуля в какой-то точке x, то он и в окрестности этой точки тоже отличен от нуля.

Естественно, возрасти $\operatorname{rang} f(x)$ может. В следующей теореме мы предполагаем, что ранг дифференциала постоянен в окрестности точки x.

Теорема о ранге. *Пусть $U \subset \mathbb{R}^m$ — окрестность точки $x_0 \in \mathbb{R}^m$. Пусть $f : U \to \mathbb{R}^n$, $f \in C^p$, $p \geqslant 1$ и $\operatorname{rang} f(x) = k$ при всех $x \in U$. Тогда существуют такие окрестности $O(x_0)$ и $O(y_0)$ точек x_0 и $y_0 = f(x_0) \in \mathbb{R}^n$ и такие диффеоморфизмы $u = \varphi(x)$, $v = \psi(y)$ класса C^p, что в окрестности $O(u_0) = \varphi(O(x_0))$ точки $u_0 = \varphi(x_0)$ отображение $v = \psi\Big(f\big(\varphi^{-1}(u)\big)\Big)$ имеет вид*

$$(u_1, \ldots, u_k, \ldots, u_m) = u \mapsto v = (u_1, \ldots, u_k, 0, \ldots, 0).$$

Для линейных отображений у вас была соответствующая теорема о том, что всегда линейное отображение можно было привести к каноническому виду, теорема о ранге — нелинейный локальный аналог.

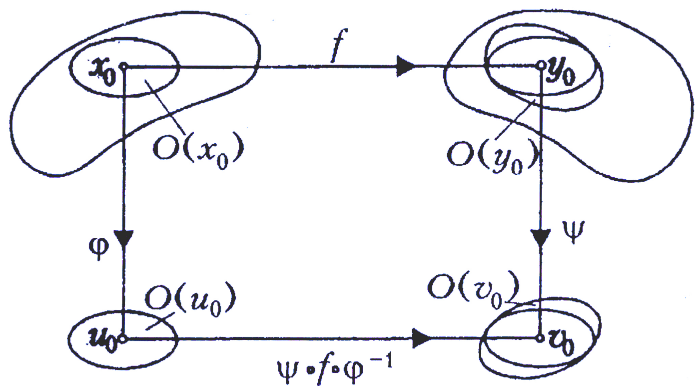
Эта теорема утверждает, что всегда можно выбрать новые координаты в области определения и в образе так, чтобы отображение ранга k имело канонический вид.

Иными словами, все такие отображения разбиваются на классы по инварианту k.

Доказательство. Обозначения. Без ограничения общности можно считать, что отображение $y = f(x)$ устроено так. Вектор $x \in \mathbb{R}^m$ имеет 2 части: $x^1 = (x_1, \dots, x_k)$ и $x^2 = (x_{k+1}, \dots, x_m)$, вектор $y \in \mathbb{R}^n$ также имеет 2 части $y^1 = (x_1, \dots, x_k)$ и $y^2 = (y_{k+1}, \dots, y_n)$. То, что $m, n \geqslant k$, очевидно. Отображение f теперь будем записывать в виде $y^1 = f^1(x^1, x^2),\ y^2 = f^2(x^1, x^2)$.

То есть все переменные и функции с индексами сверху — это векторы, с индексом 1 — векторы размерности k, с индексом 2 — векторы размерности $m - k$ или $n - k$, по смыслу.

При этом мы полагаем, что именно дифференциал $(y^1)'_{x^1}$ обратим. Иными словами, именно компоненте y^1 соответствует тот минор с ненулевым определителем.



Рассмотрим отображение $u^1 = f^1(x^1, x^2),\ u^2 = x^2$ в $\mathbb{R}^m$. Его матрица Якоби имеет вид

$$\begin{pmatrix} (y^1)'_{x^1} & (y^1)'_{x^2} \\ 0 & Id \end{pmatrix}.$$

Это я так записал блочную матрицу размера $m \times m$, у нее блок $(y^1)'_{x^1}$ имеет размер $k \times k$, блок Id имеет размер $(m-k) \times (m-k)$, блок $(y^1)'_{x^2}$ имеет размер $k \times (m-k)$. По предположению, это обратимая матрица.

Значит, по теореме о диффеоморфизме, отображение $u = \varphi(x)$ является C^p-диффеоморфизмом в некоторой окрестности точки x_0, определено обратное отображение $\varphi^{-1} \in C^p$.

По построению, композиция $g = f(\varphi^{-1}(u))$ определена в окрестности окрестности точки u_0 и имеет представление

$$y^1 = f^1(\varphi^{-1}(u)) = u^1, \qquad y^2 = f^2(\varphi^{-1}(u)) = g(u^1, u^2).$$

Посчитаем матрицу Якоби отображения g:

$$\begin{pmatrix} Id & 0 \\ g'_{u^1} & g'_{u^2} \end{pmatrix}.$$

А теперь вспомним, что отображение g не может иметь ранг больше k, так как его дифференциал есть произведение матрицы полного ранга на матрицу ранга k. Значит, $g'_{u^2} = 0$, то есть $g = (g^1, g^2)$, $g^1 = Id$ и компонента g^2 не зависит от переменной u^2, поэтому $g^2 = g(u_1)$.

Значит отображение $g = f(\varphi^{-1}(u))$ можно переписать в виде $y^1 = u^1,\ y^2 = g(u^1)$.

Теперь положим $v = (v^1, v^2) \in \mathbb{R}^n$, $v = \psi(y)$, $v^1 = \psi^1(y) = y^1,\ v^2 = \psi^2(y) = y^2 - g(y^1)$. Это отображение в $\mathbb{R}^n$, класса C^p, его матрица Якоби имеет вид

$$\begin{pmatrix} Id & 0 \\ g'_{y^1} & Id \end{pmatrix}.$$

Его определитель равен 1, значит это диффеоморфизм.

Диффеоморфизмы φ и ψ — это как раз те, о которых говорилось в формулировке теоремы.

Нужно еще только проследить за окрестностями. Мы 2 раза использовали теорему о диффеоморфизме, там каждый раз выбирались окрестности, вот каждый раз надо брать пересечения всех этих окрестностей. $\square$

Замечания к теореме о ранге.

1. Доказали теорему для случая $k < n, m$. Случаи $k = n$ и $k = m$ доказываются точно так же, только проще.

2. Если ранг отображения f равен n, то точка $y = f(x)$ является внутренней.

Это просто: $f = \psi^{-1} \circ \Big(\psi \circ f \circ \varphi^{-1}\Big) \circ \varphi$, по краям стоят диффеоморфизмы, переводящие внутренние точки во внутренние, а в середине стоит отображение $u \mapsto v$.

3. Если ранг отображения f равен k и $k < n$, то при правильном разбиении f на компоненты в некоторой окрестности точки x имеют место соотношения $f^2(x) = g^2(f^1(x))$. Мы это по дороге использовали.

Зависимость функций.

Определение. Система непрерывных функций $f^k,\ k = 1, \ldots, n,\ f^k : \mathbb{R}^m \to \mathbb{R}$ называется *функционально независимой в окрестности точки* $x_0 \in \mathbb{R}^m$, если для любой непрерывной функции $F : \mathbb{R}^n \to \mathbb{R}$, определенной в окрестности точки $y_0 = (f^1(x_0), \ldots, f^1(x_0)) \in \mathbb{R}^n$, соотношение

$$F(f^1(x), \ldots, f^n(x)) \equiv 0$$

возможно только для $F \equiv 0$.

Если система не является функционально независимой в окрестности точки $x_0 \in \mathbb{R}^m$, то она называется *функционально зависимой*.

Линейная зависимость-независимость — близкое понятие. Только функции все линейные.

Пусть дана система гладких функций $f^j, \ j = 1, \ldots, n, \ f^j : \mathbb{R}^m \to \mathbb{R}$. В следующем утверждении мы предполагаем, что ранг дифференциала отображения $f : \mathbb{R}^m \to \mathbb{R}^n$, $f(x) = (f^1(x), \ldots, f^n(x))$ постоянен в окрестности точки x_0 и равен k.

Утверждение.

а) *При $k = n$ система гладких функций f^j является функционально независимой.*

б) *При $k < n$ система функций f^j является функционально зависимой. При этом можно выбрать такие k индексов s_j, что система функций $f^{s_j}, \ j = 1, \ldots, k$ функционально независима и остальные функции $f^s, \ s \neq s_j$ можно записать в виде*

$$f^s(x) = g^s(f^{s_1}(x), \ldots, f^{s_k}(x)),$$

где g^s — гладкие функции, определенные в окрестности точки y_0 и зависящие только от k координат с индексами s_j.

Эту теорему легко запомнить так: вспомнить про линейную зависимость, здесь все то же самое: есть n векторов размерности m, из их координат составим матрицу, получим матрицу, если ранг k равен n, то векторы линейно независимы, если $k < n$, то векторы линейно зависимы.

Доказательство этой теоремы легко следует из сделанных замечаний к теореме о ранге.

При $k = n$ при отображении $y = f(x)$ образ окрестности точки x_0 содержит целую окрестность точки y_0. Значит, соотношение

$$F(f^1(x), \ldots, f^n(x)) \equiv 0$$

эквивалентно соотношению $F(y) \equiv 0$ в этой окрестности.

Теперь пусть $k < n$ и ранг k реализуется на первых k функциях f_i. Тогда в силу теоремы о ранге (в силу замечания 2) найдутся $n-k > 0$ функций g_i таких, что $y_i = g_i(y_1, \ldots, y_k)$, каждую из которых можно взять в качестве F. $\square$

Лекция 12 (27 мая 2016)

На прошлой лекции я немножко поговорил о диффеоморфизмах, в качестве приложения теоремы о диффеоморфизме я «распрямил кривую». Основная часть лекции была про теорему о ранге. И в конце я начал рассказ о разложении диффеоморфизма на простейшие. И довольно убедительно сказал, почему линейный диффеоморфизм раскладывается на простейшие.

Разложение диффеоморфизма в композицию простейших.

Определение. Диффеоморфизм $f : U \subset \mathbb{R}^m \to \mathbb{R}^m$ назовем простейшим, если если его координатное представление имеет при некотором j вид

$$f_i(x) = x_i, \quad i \neq j, \qquad f_j(x) \text{ — какая-то функция.}$$

Иначе, простейший диффеоморфизм меняет только одну координату диффеоморфизма.

Теорема. *Всякий диффеоморфизм в окрестности точки может быть представлен в виде композиции простейших диффеоморфизмов.*

Первое соображение. Попробуем представить себе линейные диффеоморфизмы, то есть квадратные невырожденные матрицы. Верно ли это для таких матриц?

Я не помню уже, проходят ли прямо вот такую теорему в обычном курсе линейной алгебры. Но я знаю точно, что всякую квадратную матрицу можно можно привести к треугольному виду элементарными преобразованиями, это метод Гаусса решения систем линейных уравнений. Каждое элементарное преобразование — простейший диффеоморфизм. А треугольную матрицу легко представить в виде композиции простейших: преобразование

$$y_1 = a_{11}x_1, \qquad y_2 = a_{21}x_1 + a_{22}x_2, \qquad y_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3, \qquad \dots$$

очевидно, является произведением простейших: сначала преобразовываем n-ю координату, потом $n-1$-ю координату и так далее.

Второе соображение. Таким образом, преобразование, переставляющие координаты тоже является композицией простейших. Ну это можно и «в лоб» доказать, композицией трёх простейших:

$$\begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Доказательство этой теоремы я буду производить в координатах и по индукции. Я думал, как записать доказательство в некоординатной форме и без индукции, и у меня внятно не вышло. Поэтому я буду доказывать все в координатах, это будет довольно громоздко, но, надеюсь, понятно.

Предположение индукции: любой диффеоморфизм, меняющий не более k координат представим в виде композиции простейших.

База индукции: при $k = 1$ утверждение очевидно — такой диффеоморфизм простейший.

Пусть любой диффеоморфизм, меняющий не более $k-1$ координат представим в виде композиции простейших. Рассмотрим диффеоморфизм, меняющий не более k координат.

Рассмотрим диффеоморфизм, меняющий первые k координат:

$$\begin{aligned} y_1 &= f_1(x_1, \dots, x_m), \\ &\dots\dots\dots \\ y_k &= f_k(x_1, \dots, x_m), \\ y_{k+1} &= x_{k+1}, \\ &\dots\dots\dots \\ y_m &= x_m. \end{aligned}$$

Первые координаты мы выбираем не меняя общности в силу сделанных сначала замечаний про линейные диффеоморфизмы.

Поскольку f — диффеоморфизм, то его матрица Якоби не вырожденная, имеет обратную. Зафиксируем некоторую точку x_0, напишем определитель матрицы Якоби

$$\begin{vmatrix} \partial_1 f_1 & \dots & \partial_k f_1 & \vdots & \partial_{k+1} f_1 & \dots & \partial_m f_1 \\ \dots & \dots & \dots & \vdots & \dots & \dots & \dots \\ \partial_1 f_k & \dots & \partial_k f_k & \vdots & \partial_{k+1} f_k & \dots & \partial_m f_k \\ \dots & \dots & \dots & \vdots & \dots & \dots & \dots \\ & & & \vdots & 1 & \dots & 0 \\ & 0 & & \vdots & & & \\ & & & \vdots & 0 & \dots & 1 \end{vmatrix} = \begin{vmatrix} \partial_1 f_1 & \dots & \partial_k f_1 \\ \dots & \dots & \dots \\ \partial_1 f_k & \dots & \partial_k f_k \end{vmatrix} \neq 0.$$

Из последнего соотношения можно сделать вывод, что один из миноров размера $k-1$ отличен от нуля. Можно считать, что это главный минор, первые $k-1$ функций, первые $k-1$ переменных. Теперь рассмотрим вспомогательное отображение $g : U \to \mathbb{R}^m$, определяемое равенствами

$$\begin{aligned} u_1 &= f_1(x_1, \dots, x_m), \\ &\dots\dots\dots \\ u_{k-1} &= f_{k-1}(x_1, \dots, x_m), \\ u_k &= x_k, \\ &\dots\dots\dots \\ u_m &= x_m. \end{aligned}$$

Якобиан этого отображения имеет вид

$$\begin{vmatrix} \partial_1 f_1 & \dots & \partial_{k-1} f_1 & \vdots & \partial_k f_1 & \dots & \partial_m f_1 \\ \dots & \dots & \dots & \vdots & \dots & \dots & \dots \\ \partial_1 f_{k-1} & \dots & \partial_{k-1} f_{k-1} & \vdots & \partial_k f_{k-1} & \dots & \partial_m f_{k-1} \\ \dots & \dots & \dots & \vdots & \dots & \dots & \dots \\ & & & \vdots & 1 & \dots & 0 \\ & 0 & & \vdots & & & \\ & & & \vdots & 0 & \dots & 1 \end{vmatrix} = \begin{vmatrix} \partial_1 f_1 & \dots & \partial_{k-1} f_1 \\ \dots & \dots & \dots \\ \partial_1 f_k & \dots & \partial_{k-1} f_{k-1} \end{vmatrix} \neq 0$$

(по предположению). Значит, g — диффеоморфизм и есть обратный диффеоморфизм g^{-1}.

Пусть $h = f(g^{-1})$. Отображение h переводит u в x, потом f переводит x в y.

Отображение h — это композиция диффеоморфизмов, следовательно h — тоже диффеоморфизм. Его координатная запись имеет вид

$$\begin{aligned} y_1 &= f_1(g^{-1}(u)) = u_1, \\ &\dots\dots\dots \\ y_{k-1} &= f_{k-1}(g^{-1}(u)) = u_{k-1}, \\ y_k &= f_k(g^{-1}(u)), \\ y_{k+1} &= u_{k+1} \\ &\dots\dots\dots \\ y_m &= u_m, \end{aligned}$$

то есть h — это простейший диффеоморфизм.

Теперь всё просто: $f(x) = h(g(x))$, по предположению индукции g раскладывается в композицию простейших, значит и f раскладывается в композицию простейших. $\square$

Теперь можно еще заметить, что по доказательству кажется, что простейших диффеоморфизмов ровно m штук: сколько-то было в разложении h и еще g. Это не так уже в линейном случае. Где-то там в середине шага индукции у нас были перестановки переменных, надо их тоже посчитать.

Теперь мы переходим к следующей теме, немножко в сторону, мы говорили про экстремумы, мы говорили про диффеоморфизмы, сейчас я займусь объединяющей все это темой.

Лемма Морса.

Пусть дана функция $f : U \in \mathbb{R}^m \to \mathbb{R}$, $f \in C^2$. Пусть $x_0 \in U$ — критическая точка этой функции, то есть $f'(x_0) = 0$, или $\nabla f(x_0) = 0$, или $\partial_k f(x_0) = 0$ — это все одно и тоже.

Определение. Критическая точка называется *невырожденной*, если матрица

$$\begin{pmatrix} \partial_{11}f & \dots & \partial_{1m}f \\ \dots & \dots & \dots \\ \partial_{m1}f & \dots & \partial_{mm}f \end{pmatrix}$$

(***гессиан*** функции f) невырожденная, то есть у нее не равен нулю определитель.

Если 0 — критическая точка достаточно гладкой функции, то по формуле Тейлора

$$f(x) - f(0) = \frac{1}{2} \sum \partial_{ij} f(0) x_i x_j + o(\|x\|).$$

Лемма Морса утверждает, что всегда существует такой диффеоморфизм окрестности нуля в окрестность нуля, что в новых координатах $u = (u_1, \dots, u_m)$ функция f будет иметь вид

$$f(u) - f(0) = -u_1^2 - \dots - u_k^2 + u_{k+1}^2 + \dots + u_m^2.$$

Если бы в формуле Тейлора не было малых слагаемых, то $f(x) - f(0)$ была бы квадратичная форма, как известно из линейной алгебры, ее можно было бы привести к указанному виду линейным преобразованием. В общем случае это можно сделать нелинейным диффеоморфизмом, причем только локально.

Для доказательства сначала сформулируем и докажем лемму Адамара.

Лемма Адамара. *Пусть функция $f : U \to \mathbb{R} \in C^p$, $p \geqslant 1$ определена в выпуклой окрестности $U \subset \mathbb{R}^m$ точки $0 \in \mathbb{R}^m$. Пусть $f(0) = 0$. Тогда существует отображение $g(x) = (g_1(x_0, \dots, g_m(x)) \in C^{p-1}$ такие, что справедливо равенство $f(x) = (x, g(x))$ или, что тоже, $f(x_1, \dots, x_m) = \sum\limits_{i=1}^{m} x_i g_i(x_1, \dots, x_m)$, причем $g_i(0) = \partial_i f(0)$.*

Доказательство. Требуемая формула вытекает из равенства

$$f(x) = \int_0^1 \frac{df(tx)}{dt}\, dt = \sum_{i=1}^{m} x_i \int_0^1 \partial_i f(tx)\, dt,$$

если положить

$$g_i(x) = \int_0^1 \partial_i f(tx)\, dt, \qquad i = 1, \dots, m.$$

Равенство $g_i(0) = \partial_i f(0)$ следует из формулы и непрерывности g, то, что $g \in C^{p-1}$, — очевидно. $\square$

Теперь перейдем к точной формулировке и доказательству леммы Морса.

Лемма Морса. *Пусть $f : G \to \mathbb{R}$, $f \in C^3$, пусть $x_0 \in G$ — внутренняя точка G является невырожденной критической точкой f. Найдется такой диффеоморфизм $g : V \to U$, окрестности V окрестности нуля на окрестность $U \subset G$, что*

$$f(g(u)) = f(x_0) - u_1^2 - \dots - u_k^2 + u_{k+1}^2 + \dots + u_m^2.$$

Доказательство. Без ограничения общности и линейными заменами переменных можно свести задачу к случаю $x_0 = 0$ и $f(x_0) = 0$.

Применим к функции f лемму Адамара, получим $f = (x, g(x))$. Так как 0 — критическая точка, то $g(0) = 0$ в силу той же леммы. Теперь применим ко всем координатам g_i ту же лемму Адамара:

$$g_i(x) = (x, h_i(x)) = \sum_{j=1}^{m} x_j h_{ij}(x).$$

Таким образом

$$f(x) = \sum_{i,j=1}^{m} x_i x_j h_{ij}(x).$$

Такое представление не единственно, для функций h_{ij} и h_{ji} фиксирована их сумма. Без ограничения общности можно считать, что $h_{ij} = h_{ji}$.

Из непрерывности h_{ij} и единственности тейлоровского разложения следует равенство $h_{ij}(0) = \partial_{ij} f(0)$. По предположению матрица $h_{ij}(0)$ не вырождена.

Теперь функция f записана в матричном виде (что-то вроде билинейной формы), а мы будем приводить её диффеоморфизмами к диагональному виду. Действуем по индукции.

Предположение индукции. Существуют координаты $u = (u_1, \ldots, u_m)$ (иными словами — диффеоморфизм $x = \varphi(u)$, $\varphi(0) = 0$), при которых

$$f(\varphi(u)) = \pm u_1^2 \pm \ldots \pm u_{r-1}^2 + \sum_{i,j=r}^{m} u_i u_j H_{ij}(u).$$

Здесь $H_{ij} = H_{ji}$ — некоторые функции. При $r = 1$ равенство уже доказано с применением леммы Адамара.

Предположим, что при $r > 1$ это равенство верно, докажем его при $r + 1$.

Так как матрица $h_{ij}(0)$ не вырождена и φ — диффеоморфизм, то матрица формы

$$\pm u_1^2 \pm \ldots \pm u_{r-1}^2 + \sum_{i,j=r}^{m} u_i u_j H_{ij}(0)$$

также не вырождена[4]. Линейной заменой переменных u_j, $j = r, \ldots, m$ всегда можно привести форму к диагональному виду, поэтому можно считать, что в предыдущей формуле $H_{rr}(0) \neq 0$. По непрерывности $H_{rr}(x) \neq 0$ в некоторой окрестности нуля.

[4]Специально для Пети: есть функция $w(x) = (Ax, x) + O(\|x\|^3)$, причём A — невырожденная матрица, она же гессиан w по x. Теперь пусть $x = \varphi(u)$, где $\varphi(u) = Cu + O(\|u\|^2)$ — диффеоморфизм в окрестности 0. Тогда C — невырожденная и если собрать эти 2 формулы получится, что $w(u) = (C^T ACu, u) + O(\|u\|^3)$ в нуле, то есть гессиан w по u равен $C^T AC$ и тоже не вырожден.

Положим $\psi(u) = \sqrt{|H_{rr}(u)|}$. Функция ψ непрерывно дифференцируема в некоторой окрестности нуля, сделаем теперь переход к новым координатам

$$v_i = u_i, \qquad i \neq r$$
$$v_r = \psi(u)\left(u_r + \sum_{j>r} \frac{u_j H_{jr}(u)}{H_{rr}(u)}\right).$$

Матрица Якоби (в нуле!) преобразования $v = \phi(u)$ отличается от единичной одной строкой номер r, в этой строке при $u = 0$ на диагонали стоит число $\psi(0)$. Якобиан $\psi(0) \neq 0$, поэтому отображение ϕ — диффеоморфизм.

Теперь запишем основное равенство для $f(\varphi(u))$ в переменных v. Для этого сначала в слагаемых $\sum_{i,j=r}^m u_i u_j H_{ij}(u)$ выделим слагаемые, содержащие u_r:

$$B = u_r^2 H_{rr}(u) + 2u_r \sum_{j=r+1}^m u_j H_{rj}(u)$$

и запишем их в переменных v: $B = \pm v_r^2 - \dfrac{1}{H_{rr}(v)} \displaystyle\sum_{i,j=r+1}^m v_i v_j H_{ij}(v)$. Знак $\pm$ перед v_r^2 совпадает со знаком $H_{rr}(0)$. Таким образом шаг индукции полностью сделан. □

Замечания к лемме Морса.

1. Из леммы Морса следует, что все невырожденные критические точки являются изолированными. При диффеоморфизмах $u = \varphi(x)$ критические точки u функции $f(u)$ однозначно соответствуют критическим точкам x функции $f(\varphi(x))$ переходят в критические $(\big(f(\varphi(x))\big)' = \varphi'(x) f'(\varphi(x)))$.

У функции $f(x_0) - u_1^2 - \ldots - u_k^2 + u_{k+1}^2 + \ldots + u_m^2$ в окрестности 0 (и не только) есть единственная критическая точка 0, она изолированная, значит и x_0 — невырожденная критическая точка.

2. Как и в линейном случае, справедлив закон инерции: количество знаков + и знаков − в каноническом представлении не зависит от способа приведения к каноническому виду. Иными словами, эти количества являются свойством функции и невырожденной критической точки. Они полностью определяются гессианом функции. Число знаков минус называется *индекс критической точки.* У локального минимума индекс равен нулю, у локального максимума индекс равен m.

3. Заметьте, что в одномерном случае критическая точка x невырождена, если $f''(x) \neq 0$. Каждая невырожденная критическая точка является точкой экстремума.

В двумерном случае бывают 3 типа невырожденных критических точек: максимум, минимум и обычное седло. Обезьянье седло — вырожденная критическая точка.

Лекция 13 (30 мая 2016)

На прошлой лекции мы локально разложили диффеоморфизм в произведение простейших. Кроме того мы обсудили и доказали лемму Морса, о том, что если функция f в окрестности точки (критической) имеет вид «квадратичные слагаемые» + «старшие нелинейности», то есть правильные координаты, в которых f имеет канонический вид $\sum \pm u_j^2$.

Поверхности.

Вам будут рассказывать про то, что такое поверхность, много раз и на разных дисциплинах. Мы представляем себе обычно двумерную поверхность в трёхмерном пространстве и как-то переносим такую объёмную картинку в многомерные пространства. Кривая — это простейший вид поверхности: поверхность размерности 1.

Понятно, что как геометрический объект мы представляем себе всё только в трёхмерном пространстве. Решать задачи приходится в многомерном пространстве. Поэтому такую важную роль играют аналитические методы изучения кривых поверхностей. Простейший случай вы изучали — это квадрики.

Вроде очевидно, что кривые и поверхности можно задавать различными формулами. В смысле, что одну и ту же поверхность можно задавать по-разному, различными формулами.

Основных способов 2: параметрический и неявный.

В параметрическом способе мы предполагаем, что геометрическая фигура, которая получается в результате гладкого отображения области $G \in \mathbb{R}^m$ в пространство $\mathbb{R}^n$, имеет при $m < n$ размерность m. Я не обсуждаю, что значит множество размерности m. Это сложное понятие, бывают различные определения, вы слышали слова про дробные размерности, про фракталы. Я в настоящий момент философствую и говорю об интуитивных понятиях.

Типичный пример: график функции 2х переменных в $\mathbb{R}^3$.

Другой важный способ задания поверхностей — уравнения. Сфера в $\mathbb{R}^3$ задается уравнением $x^2 + y^2 + z^2 = 1$. Кривая — это пересечение 2х поверхностей, то есть множество решений 2х уравнений. Конечно, интуитивно всегда «количество уравнений» + «размерность поверхности» = «размерность пространства».

Очень важно понимать, что локальное устройство множества и глобальное его устройство - это совершенно разные вещи. Локально каждая точка окружности и каждая точка интервала (неважно, в 2-х или 3-х –мерных пространствах) — одно и то же. Глобально мы все понимаем, что окружность и интервал — вещи разные. Что означает, слово «разные» в этом контексте, очевидно, это вопросы не матанализа, а разных других математических наук.

А граница квадрата и окружность — они разные или одинаковые? На этот вопрос ответы могут быть разные — и правильные. То, что они гомеоморфные (есть гомеоморфизм, переводящий одно в другое), легко увидеть.

Более того, не просто существует гомеоморфизм окружности и квадрата. Существует гомеоморфизм областей, содержащих окружность и квадрат.

Мы не будем заниматься глобальным устройством множеств (кривых, поверхностей). Вместо этого займемся их локальным устройством.

Основное определение. *Множество $S \subset \mathbb{R}^n$ называется* k-мерной поверхностью гладкости p, *если для любой точки $x_0 \in S$ найдется окрестность $U(x_0) \in \mathbb{R}^n$, и диффеоморфизм $\varphi : U(x_0) \mapsto \mathbb{R}^n$ гладкости p окрестности $U(x_0)$ на окрестность $O \subset \mathbb{R}^n$ начала координат, такие что множество $\varphi(S \bigcap U(x_0))$ совпадает с множеством $\{t \in O : t_{k+1} = \ldots = t_n = 0\}$.*

Это определение сразу означает, что локально есть $n-k$ уравнений «диффеоморфизм равен нулю».

Важно, что мы диффеоморфно отображаем шар в большом пространстве, не только саму поверхность, но и её окрестность.

Негладкое определение. *Поверхностью размерности k в $\mathbb{R}^n$ называется множество $S \in \mathbb{R}^n$, каждая точка которого имеет окрестность $S \bigcap \mathbb{R}^n$, гомеоморфную $\mathbb{R}^k$.*

Окрестность нуля может быть любой, например, шар, или промежуток–параллелепипед.

Вместо слова поверхность ещё используют слова многообразие или подмногообразие.

Ясно, что такое определение отличается от предыдущего. В нём нет и не используется никакой гладкости. Никак не используется как именно располагается поверхность в $\mathbb{R}^n$.

Квадрат (как множество из четырёх отрезков) не является 1-мерной поверхностью гладкости 1 в смысле первого определения, но является поверхностью гладкости 0 в смысле второго определения.

Почему квадрат (из-за угловых точек) не является гладкой поверхностью.
1. Если есть диффеоморфизм f гладкости 1 и выше, то его дифференциал обратим. Это просто: $Id = f \circ f^{-1},\ D(Id) = Id,\ D(f \circ f^{-1}) = Df \circ D(f^{-1})$, а это — произведение матриц (по правилу дифференцирования сложной функции).
2. Пусть есть угол квадрата и он диффеоморфен отрезку. То есть есть отображение $(u, v) = f(x, y)$, такое что угол «$\llcorner$» переходит в отрезок $v = 0,\ u \in \Delta$, то есть $\partial_x f(0,0) = \partial_y f(0,0) = 0$.
3. Получили противоречие — вырожденный дифференциал.

Другой пример: рогатая сфера Александера — поверхность, гомеоморфная 2-мерной сфере в $\mathbb{R}^3$ ограничивает область не гомеоморфную шару.

Теорема (параметрическое определение). *Пусть отображение окрестности* $\varphi : U \in \mathbb{R}^k \to \mathbb{R}^n,\ n > k$ *имеет гладкость* $C^p,\ p \geqslant 1$ *и пусть* $\operatorname{rang} \varphi = k$ *в каждой точке* U*. Тогда образ* $\varphi(U(t))$ *некоторой окрестности* $U(t) \subset U$ *каждой точки* $t \in U$ *— это поверхность гладкости* p *размерности* k *в* $\mathbb{R}^n$*.*

Иными словами, если отображение, задающее поверхность параметрически, гладкое и его ранг максимальный, то локально эта поверхность удовлетворяет первому определению поверхности тоже. В критических точках, где ранг меньше, может быть что угодно.

Глобально поверхность параметрически не получится задать даже с помощью отображения постоянного ранга (пример — та же сфера или окружность). Требуется дополнительное условие, например, что φ — гомеоморфизм (взаимно однозначное отображение).

Доказательство. Выберем произвольную точку $x_0 \in \varphi(U)$, пусть $x_0 = \varphi(t_0),\ t_0 \in U$.

Без ограничения общности считаем, что координаты пронумерованы так, что минор $\partial_i \varphi_j,\ i, j = 1, \ldots, k$ отличен от нуля в окрестности точки x_0. Тогда соотношения

$$
\begin{aligned}
x_1 &= \varphi_1(t_1, \ldots, t_k), \\
&\ldots\ldots\ldots \\
x_k &= \varphi_k(t_1, \ldots, t_k), \\
x_{k+1} &= \varphi_{k+1}(t_1, \ldots, t_k), \\
&\ldots\ldots\ldots \\
x_n &= \varphi_n(t_1, \ldots, t_k)
\end{aligned}
$$

в силу теоремы о неявной функции можно переписать в виде

$$
\begin{aligned}
t_1 &= f_1(x_1, \ldots, x_k), \\
&\ldots\ldots\ldots \\
t_k &= f_k(x_1, \ldots, x_k), \\
x_{k+1} &= \varphi_{k+1}(t_1, \ldots, t_k), \\
&\ldots\ldots\ldots \\
x_n &= \varphi_n(t_1, \ldots, t_k)
\end{aligned}
$$

и, следовательно, в виде

$$\begin{aligned}
t_1 &= f_1(x_1, \ldots, x_k),\\
&\cdots\cdots\cdots\\
t_k &= f_k(x_1, \ldots, x_k),\\
x_{k+1} &= f_{k+1}(x_1, \ldots, x_k) = \varphi_{k+1}(f_1(x_1, \ldots, x_k), \ldots, f_k(x_1, \ldots, x_k)),\\
&\cdots\cdots\cdots\\
x_n &= f_n(x_1, \ldots, x_k) = \varphi_n(f_1(x_1, \ldots, x_k), \ldots, f_k(x_1, \ldots, x_k)).
\end{aligned}$$

Теперь рассмотрим отображение в $\mathbb{R}^n$ окрестностей точек t_0 и x_0:

$$\begin{aligned}
t_1 &= f_1(x_1, \ldots, x_k),\\
&\cdots\cdots\cdots\\
t_k &= f_k(x_1, \ldots, x_k),\\
t_{k+1} &= x_{k+1} - f_{k+1}(x_1, \ldots, x_k),\\
&\cdots\cdots\cdots\\
t_n &= x_n - f_n(x_1, \ldots, x_k).
\end{aligned}$$

Это отображение в силу теоремы о диффеоморфизме является диффеоморфизмом окрестностей в $\mathbb{R}^n$, при этом поверхность S переходит в требуемое множество $t_{k+1} = \ldots = t_n = 0$. $\square$

Примеры.

1. Само пространство $\mathbb{R}^n$ является n-мерной поверхностью в себе.

2. Любое k-мерное линейное многообразие, подпространство или афинное, является поверхностью размерности k.

3. Путь есть кривая Γ на плоскости, заданная скалярным уравнением $F(x, y) = 0$. Пусть F — гладкая функция, пусть $F(x_0, y_0) = 0$ и пусть точка (x_0, y_0) не является критической: то есть ранг отображения $DF(x_0, y_0)$ равен 1, то есть по крайней мере одна из производных $F'_x(x_0, y_0)$ или $F'_y(x_0, y_0)$ отлична от нуля.

Мы доказали не так давно, что в некоторой окрестности O точки (x_0, y_0) кривая Γ — гладкое многообразие размерности 1.

Это было утверждение о возможности выпрямления кривой, в точности то, что нам нужно: существует диффеоморфизм окрестности O в окрестность начал координат, так что Γ отображается в ось абсцисс.

4. Теперь пусть всё то же самое, уравнение от многих переменных в пространстве $\mathbb{R}$, то есть $F : \mathbb{R}^{m+1} \to \mathbb{R}$, одно скалярное уравнение, переменных $m + 1$ — множество решений —

поверхность размерности m. Рассмотрим уравнение $F(x) = 0$. Пусть при всех $x_0 : F(x_0) = 0$ справедливо $\mathbf{grad}\, F(x_0) \neq 0$ (то есть хотя бы одна из производных $\partial_i F \neq 0$).

Мы доказали не так давно, что это есть гладкая поверхность размерности m в $\mathbb{R}^{m+1}$.

5. График гладкой функции $f : \mathbb{R}^m \to \mathbb{R}$ является гладкой поверхностью размерности m в $\mathbb{R}^{m+1}$. Диффеоморфизм зададим формулами

$$t_i = x_i, \quad i = 1, \ldots, m-1; \qquad t_{m+1} = x_{m+1} - f(x_1, \ldots, x_m).$$

6. Окружность S на плоскости есть гладкая поверхность размерности 1 в $\mathbb{R}^2$.

Для того, чтобы это доказать явно укажем диффеоморфизм. Вспомним полярную систему координат, положим $x = (r+1)\cos\varphi,\ y = (r+1)\sin\varphi$. Точка с координатами (x_0, y_0) перейдет в точку с координатами $r = 0,\ \varphi = \varphi_0$, окружность S локально перейдет в прямую $r = 0$.

7. Теперь самый главный, «настоящий» пример.

Пусть есть гладкое отображение $F : \mathbb{R}^n \to \mathbb{R}^{n-k}$ постоянного ранга $n-k$. Тогда уравнение $F(x) = 0$ задает в $\mathbb{R}^n$ поверхность размерности k.

В координатах уравнение имеет привычный вид

$$\begin{cases} F_1(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) = 0, \\ \qquad\ldots \qquad \ldots \\ F_{n-k}(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) = 0. \end{cases}$$

Здесь $x^1 = (x_1, \ldots, x_k),\ x^2 = (x_{k+1}, \ldots, x_n)$

Чтобы это доказать, надо написать уравнение $F(x^1, x^2) = 0,\ x^1 \in \mathbb{R}^k, x^2 \in \mathbb{R}^{m-k}$, считать, что дифференциал $\partial_{x_2} F(x_0)$ обратим, а потом сказать, что по теореме о неявной функции исходное уравнение можно переписать в виде $x^2 = \Phi(x^1)$. Поэтому можно написать диффеоморфизм $(y^1, y^2) = \varphi(x^1, x^2) = (x^1, x^2 - \Phi(x^1))$, это диффеоморфизм так как у него $D\Phi = Id$, решения уравнения $F(x) = 0$ переходят в требуемое множество $\{y^2 = 0\}$.

Более того, **всякую гладкую поверхность можно задать в виде уравнения.** В самом деле, по определению существует диффеоморфизм этой поверхности на множество $\{t \in O : t_{k+1} = \ldots = t_m = 0\}$. Вот уравнения и написаны — локальные!

Этот пример показывает, что не случайно во всех или почти всех конструкциях и теоремах задание поверхности в виде уравнений — это основной способ.

8. Рассмотрим уравнение $xy = 0$. В нуле у него градиент равен нулю. И это уравнение — из-за начала координат — не задает поверхности размерности 1 (это оси координат).

Всё дело в том, в начале координат ранг производной равен не 1, а равен 0. Предыдущая конструкция не применима.

Если ноль выкинуть, то останется 4 луча, каждый их которых является поверхностью.

9. Замкнутый луч на плоскости не является поверхностью размерности 1: в крайней точке неприятности. В $\mathbb{R}^3$ — замкнутая полуплоскость не является поверхностью размерности 2. Такое называется *поверхностью с краем.*

10. Параметрический способ задания поверхности был локальный! Окрестность каждой точки из множества параметров была поверхностью, которую можно было задать уравнениями. Глобально множество, заданное параметрически может оказаться плохим.

Достаточно, чтобы φ был гомеоморфизмом.

Пусть V — окрестность t_0, образ которой является гладкой поверхностью в $\mathbb{R}^n$. Нам достаточно найти окрестность $D \subset \mathbb{R}^n$ точки $x_0 = \varphi(t_0) \in \mathbb{R}^n$, которая не пересекается с $\varphi(U) \setminus \varphi(V)$.

Множество $\varphi(V)$ открыто в $\varphi(U)$ (т.к. φ — гомеоморфизм на свой образ), поэтому каждая точка $x \in \varphi(V)$ имеет окрестность $D_x \subset \mathbb{R}^n$, не пересекающуюся с $\varphi(U \setminus V)$, объединение таких окрестностей D_x есть такое открытое множество D в R^n, что $D \cap \varphi(U) = \varphi(V)$. Ясно, что D — искомое.

Лекция 14 (06 июня 2016)

На прошлой лекции мы ввели понятие поверхности. Дали определение, дали два варианта конструкции поверхности.

Неявно, в виде системы уравнений, $m+n$ уравнений, n неизвестных, размерность поверхности m.

Параметрически. Есть отображение $\mathbb{R}^m \to \mathbb{R}^n$, $n > m$, постоянного ранга m. Тогда образ окрестности — поверхность размерности m.

Гладкость поверхности определяется естественным образом.

На этой лекции я расскажу несколько естественных конструкций и перейду к задаче об условном экстремуме.

Следующая лекция через неделю - последняя и сокращённая.

Касательные векторы и касательные пространства.

Пусть есть 2-мерная поверхность $S \subset \mathbb{R}^3$, заданная равенством $z = f(x,y)$, где f — гладкая функция. В точке $M = (x_0, y_0, z_0) \in S,\ z_0 = f(x_0, y_0)$ можно написать формулу Тейлора

$$z - z_0 = f'_x(x_0,y_0)(x-x_0) + f'_y(x_0,y_0)(y-y_0) + \text{«малые члены»}.$$

Если малые члены отбросить, то останется равенство

$$z - z_0 = f'_x(x_0,y_0)(x-x_0) + f'_y(x_0,y_0)(y-y_0),$$

геометрически оно задаёт аффинную плоскость, проходящую через M — касательную плоскость.

Если задать на этой плоскости структуру линейного пространства, поместив M в начало координат, то мы получим касательное пространство.

Касательная плоскость к поверхности в некоторой точке — это геометрический объект в большом объемлющем пространстве.

Касательное пространство — это линейное пространство размерности как поверхность.

Если 2-мерная поверхность задавалась в $\mathbb{R}^3$ уравнением $F(x,y,z) = 0$, то уравнение касательной плоскости в точке (x_0, y_0, z_0) имеет вид

$$F'_x(x_0,y_0,z_0)(x-x_0) + F'_y(x_0,y_0,z_0)(y-y_0) + F'_z(x_0,y_0,z_0)(z-z_0) = 0.$$

Пусть в $\mathbb{R}^m$ есть поверхность S, заданная уравнением $f(x) = c$ — поверхность уровня функции $f \in C^1$. Пусть $f(x_0) = c$, причём пусть градиент ∇f функции f отличен от 0, точка x_0 не критическая. Тогда поверхность уровня в окрестности точки x_0 — это гладкая поверхность. В точке x_0 можно написать график касательной гиперплоскости: $(\nabla f(x_0), x - x_0) = 0$.

Касательное пространство в точке x_0 задаётся уравнением $(\nabla f(x_0), h) = 0$: ему принадлежат все векторы h, удовлетворяющие этому равенству.

В общем случае формулы имеют примерно такой же вид.

Пусть у нас есть k-мерная поверхность $S \subset \mathbb{R}^n$, и в окрестности точки $x_0 \in S$ она задана с помощью гладкого отображения $t \in \mathbb{R}^k \mapsto x = \varphi(t) \in \mathbb{R}^n$, $x_0 = \varphi(0)$ — параметрически. Предположим, что дифференциал $D\varphi(0)$ имеет полный ранг k (тогда и в окрестности $\operatorname{rang} D\varphi(t) = k$).

Тогда k-мерная плоскость в $\mathbb{R}^n$, задаваемая параметрически формулой $x - x_0 = Dx(0)t$, называется касательной плоскостью.

В координатах формула $x - x_0 = Dx(0)t$ имеет вид

$$\begin{cases} x_1 - x_1^0 & = & \frac{\partial x_1}{\partial t_1}(0)t_1 & + \ldots + & \frac{\partial x_1}{\partial t_k}(0)t_k, \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_n - x_n^0 & = & \frac{\partial x_n}{\partial t_1}(0)t_1 & + \ldots + & \frac{\partial x_n}{\partial t_k}(0)t_k. \end{cases}$$

Если на касательной плоскости в точку касания поместить начало координат мы получим касательное пространство, естественно, оно имеет размерность k.

Пусть поверхность S в окрестности точки $x_0 \in S \subset \mathbb{R}^n$ задана уравнением $F(x) = 0$, где гладкое отображение $F : \mathbb{R}^n \to \mathbb{R}^{n-k}$ имеет ранг $n-k$. Когда я говорил о примерах поверхностей я приводил эту ситуацию в качестве главного примера.

Напишем уравнение касательной плоскости (может быть лучше говорить «гиперплоскости»).

С одной стороны, мы понимаем, что это уравнение должно иметь вид $DF(x_0)(x - x_0) = 0$ — это просто линеаризация, главное приближение уравнения $F(x) = 0$.

Чтобы привести это понимание к определению, для определенности будем считать, что $x = (u, v)$, $x \in \mathbb{R}^k$, $v \in \mathbb{R}^{n-k}$ и перепишем уравнение $F(x) = 0$ в виде $F(u, v) = 0$. Условие о ранге означает обратимость линейного отображения $D_v F(x_0)$. По теореме о неявной функции локально уравнение $F(u, v) = 0$ имеет вид $v = f(u)$.

Теперь по определению получаем параметрическое представление касательной плоскости: $u - u_0 = Id\ t,\ v - v_0 = f'(u_0)t$.

По формуле для производной из теоремы о неявной функции имеем

$$f'(u_0) = -\big(F'_v(u_0, v_0)\big)^{-1} F'_u(u_0, v_0),$$

отсюда следует формула $F'_u(u_0, v_0)(u - u_0) + F'_v(u_0, v_0)(v - v_0) = 0$, после возвращения к переменной x имеем требуемую формулу $DF(x_0)(x - x_0) = 0$, определяющую точки x касательной плоскости.

В координатах уравнения касательной плоскости имеют вид

$$\begin{cases} \frac{\partial F_1}{\partial x_1}(x_0)(x_1 - x_1^0) & + \ldots + & \frac{\partial F_1}{\partial x_n}(x_0)(x_n - x_n^0) & = 0, \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial F_{n-k}}{\partial x_1}(x_0)(x_1 - x_1^0) & + \ldots + & \frac{\partial F_{n-k}}{\partial x_n}(x_0)(x_n - x_n^0) & = 0. \end{cases}$$

Касательный вектор.

Есть гладкое отображение $x : [0, 1] \to \mathbb{R}^n$. Пусть это отображение не вырождено: в каждой точке $x' \neq 0$.

Каждое такое отображение задаёт в пространстве $\mathbb{R}^n$ кривую Γ. Прямая $x = x_0 + x'(t_0)\tau,\ \tau \in \mathbb{R}$ является касательной к кривой Γ в точке $x_0 = x(t_0)$.

Вектор с началом в точке $x_0 = x(t_0)$ и концом на касательной прямой называется касательным вектором к кривой Γ в точке $x_0 = x(t_0)$.

Теорема. *Пространство, касательное к поверхности S состоит их векторов, касательных в точке x_0 к кривым, лежащим на S и проходящих через точку x_0, и только из таких векторов.*

Вообще говоря, интуитивно (в $\mathbb{R}^3$) эта теорема очевидна.

Для **доказательства** теоремы надо доказать 2 утверждения:

1) Каждый касательный вектор принадлежит касательному подпространству;

2) Каждый элемент касательного пространства является касательным вектором к некоторой кривой на поверхности.

По определению поверхности (существует диффеоморфизм ...) её можно задать в виде уравнения $F(x) = 0$, $\dim x = n$, $\dim F = n - k$. Пусть кривая Γ задана функцией $x(t)$, причем $F(x(t)) \equiv 0$, $x(0) = x_0$.

1) Пусть кривая на S параметрически задана формулой $x(t)$. Продифференцируем тождество $F(x(t)) \equiv 0$, получаем $D\,F(x(t))x'(t) = 0$, при $t = 0$ получаем $D\,F(x_0)x'(0) = 0$, а это и есть уравнение касательного пространства.

2) Теперь выберем произвольный вектор ξ из касательного пространства к S в точке x_0, по определению $D\,F(x_0)\xi = 0$.

Необходимо указать кривую на поверхности, которая имеет вектор ξ вектором скорости.

Считаем, что поверхность задана уравнениями $F(x^1, x^2) = 0,\ x^1 \in \mathbb{R}^k, x^2 \in \mathbb{R}^{n-k},\ F \in \mathbb{R}^{n-k}$, считаем, что дифференциал $\partial_{x_2} F(x_0)$ обратим.

Тогда, зная первые k координат вектора ξ, из уравнения $D\,F(x_0)\xi = 0$ — это система линейных уравнений ранга $n - k$ — мы выразим остальные $n - k$ координат этого вектора.

Это значит, что если для вектора $(\xi^1, \tilde{\xi}^2)$ выполнено равенство $D\,F(x_0)\xi = 0$, то $\tilde{\xi}^2 = \xi^2$.

Рассмотрим прямую $x - x_0^1 = \xi^1 t$ в пространстве $\mathbb{R}^k$. Решим уравнение $F(x^1, x^2) = 0$ относительно второй переменной, получим $x^2 = f(x^1)$.

Теперь построим кривую $x = x_0^1 + \xi^1 t,\ x^2 = f(x_0^1 + \xi^1 t)$. По построению эта кривая лежит на поверхности S (так как $F(x^1, f(x^1) \equiv 0)$. Так же по построению эта кривая проходит через точку (x_0^1, x_0^2).

Теперь продифференцируем тождество $F(x_0^1 + \xi^1 t, f(x_0^1 + \xi^1 t)) = 0$ по t в точке $t = 0$ получается равенство $D_{x^1} F \xi^1 + D_{x^2} F \tilde{\xi}^2 = 0$.

Как выписать формулу для вектора $\tilde{\xi}^2$ не важно! Мы уже знаем, что $\tilde{\xi}^2 = \xi^2$. $\square$

Условный экстремум.

Постановка задачи: дана функция $f : \mathbb{R}^n \to \mathbb{R}$, дана гладкая поверхность $S \in \mathbb{R}^n$ размерности k, исследуем вопрос, когда в точке $x_0 \in S$ функция принимает экстремальное значение.

Слова «условный экстремум» означают «экстремум функции при условии $x \in S$». Функция определена везде, во всём пространстве, но добавлено условие $x \in S$.

Определения локального максимума, локального минимума, строгого локального максимума, строгого локального минимума очевидны.

Я сформулирую одно из них.

Пусть дана гладкая поверхность $S \in \mathbb{R}^n$, на ней определена вещественная функция f (определённая на поверхности, но обычно она везде определена, во всём пространстве). Точка x_0 называется точкой строго локального минимума на S, если на пересечении $S \bigcap G(x_0)$ поверхности S и некоторой окрестности $G(x_0) \subset \mathbb{R}^n$ точки x_0 справедливо соотношение

$$f(x) > f(x_0), \qquad x \in S \bigcap G(x_0).$$

Если в точке $x_0 \in S$ функция f имеет локальный экстремум во всем пространстве, то, естественно, она имеет экстремум на S. Этот случай мы считаем «не интересным». Однако может быть так, что функция имеет экстремум на S, который не является экстремумом в $\mathbb{R}^n$.

Пример. Функция $f(x, y) = x$ на окружности $x^2 + y^2 = 1$ имеет максимум в точке $x = 1, y = 0$ и минимум в точке $x = -1, y = 0$. Конечно, у этой функции нет локальных экстремумов и вообще критических точек в пространстве.

Первая мысль — считать, что поверхность задана (хотя бы локально) параметрически: $S = \{x : x = \varphi(t), t \in U \subset \mathbb{R}^k\}$, $\operatorname{rang} \varphi = k$, теперь можно переписать функцию f в виде композиции $f \circ \varphi$, и сказать, что локальный экстремум на S функции f в x_0 — это то же самое, что локальный экстремум функции $f \circ \varphi$ во всем пространстве в точке $\varphi^{-1}(x_0)$. Для него мы можем хотя бы проверить необходимое условие: надо проверить, выполнение системы k уравнений $\nabla f \circ \varphi = 0$.

Этот чудесный метод теоретически работает. Но для того, чтобы решить конкретную задачу, он совершенно не годиться.

Конечно, в указанном примере можно запараметризовать окружность $\{x = \cos t, y = \sin t\}$, и теперь у функции $f(x, y) = \cos t$ легко найти экстремумы. Однако это хорошо получается только в таких простейших случаях. Даже в случаях, когда «всё считается», правильно использовать другой метод, восходящий к Лагранжу, который я вам сейчас буду рассказывать до конца этой лекции и на следующей.

Философия и геометрия.

Я начну с визуально простого случая, когда поверхность S одномерная, то есть это кривая и лежит на плоскости. И надо найти точки локального минимума и максимума функции $f(x, y)$ на этой кривой. Каким способом задана кривая S пока не так уж и важно, например уравнением $F(x, y) = 0$.

Нарисуем картинку на плоскости и попробуем понять, как исследовать функцию f в некоторой точке (x_0, y_0) кривой на условный локальный экстремум. Условный — значит, на кривой, локальный — значит «в окрестности».

Предполагаем сразу, что точка не является критической для нашей кривой «во всем пространстве». Если она такая, то там может быть безусловный минимум. Иными словами, в этой точке хотя бы одна из частных производных $f'_x(x_0, y_0), f'_y(x_0, y_0)$ отлична от нуля, $\nabla f \neq 0$.

Я буду сейчас изображать функцию f не в виде трёхмерного графика, а в виде поверхностей уровня. Так как точка (x_0, y_0), в окрестности которой мы хотим найти экстремум, не является критической $(|f'_x(x_0, y_0)| + |f'_y(x_0, y_0)| \neq 0)$, то кривые уровня в окрестности этой точки не устроены, как циклические овалы и не являются точками пересечения. Это следует из теоремы о неявной функции: если хотя бы одна из частных производных отлична от нуля, то соответственная переменная может быть локально выражена через другую.

Поэтому мы можем считать, что линии уровня функции f — это такие «почти прямые» кривые $\Gamma_h = \{(x, y) : f(x, y) = h\}$, так сказать, параллельные друг другу и заполняющие окрестность точки (x_0, y_0). С одной стороны от кривой $\Gamma_{f(x_0,y_0)} = \Gamma$ лежат кривые с $h < f(x_0, y_0)$, с другой — кривые с $h > f(x_0, y_0)$, причём неравенства строгие.

Теперь посмотрим, как эти кривые располагаются относительно кривой S, которая проходит через точку (x_0, y_0) и имеет там общую точку с кривой Γ.

В ситуации общего положения эти две кривые (Γ и S) пересекаются под ненулевым углом. В этом случае с одной стороны от точки (x_0, y_0) на кривой S функция f принимает значения больше $f(x_0, y_0)$, с другой — меньше $f(x_0, y_0)$, поэтому в такой ситуации общего положения в точке (x_0, y_0) не может быть экстремума.

Если кривые пересекаются под нулевым углом (касаются), то экстремума тоже нет.

Геометрически единственный случай, когда экстремум возможен и действительно имеет место, это когда кривая S касается кривой Γ, причем они не пересекаются в естественном смысле.

Причем если мы по естественным соображениям умеем сказать, как расположены кривые уровня, мы можем сказать, что получится, минимум или максимум.

Вот еще геометрическая картинка.

Представим себе карту местности, гористой, и саму местность. Карта — это уменьшенная вертикальная проекция на горизонтальную плоскость. Пусть по местности идет тропа (кривая линия). Если вы идете по тропе, то вы легко увидите на ней условные экстремумы–минимумы: там будет лужица.

Ровно в этих местах на карте линии уровня будут касаться трека тропы.

Следующая геометрическая картинка в трехмерном пространстве.

Теперь у нас есть функция $f(x, y, z)$, которую мы будем представлять в виде поверхностей уровня. Представьте себе область пространства, заполненную слоистым материалом, кипу плёнок, изогнутых. На каждой пленке функция принимает некоторое постоянное значение. Снова (точка (x_0, y_0, z_0) не критическая) все эти поверхности расположены друг вдоль друга.

Пусть поверхность S — снова одномерная кривая, на этот раз в трёхмерном пространстве. Тогда из тех же построений легко увидеть, что точка (x_0, y_0, z_0) может быть условным экстремумом, только если в этой точке кривая S касается соответствующей поверхности уровня функции f.

Слова «кривая касается поверхности» означают, естественно, что вектор, касательный кривой в этой точке лежит в касательной плоскости к поверхности уровня.

Теперь мы переходим к общей теореме. Она просматривается из проведённых геометрических конструкций. Пусть в $\mathbb{R}^n$ есть точка x_0, пусть в некоторой окрестности $D \ni x_0$ задана функция $f : D \to \mathbb{R} \in C^1$, причем точка x_0 не является критической для f.

Пусть гладкая поверхность S содержит точку x_0.

Теорема (необходимый признак условного экстремума). *Для того, чтобы точка x_0 была бы точкой условного локального экстремума для функции f, необходимо выполнение условия $TS_{x_0} \subset TN_{x_0}$, где TS_{x_0} — пространство, касательное к S, а TN_{x_0} — пространство, касательное к поверхности уровня $N = \{x \in D : f(x) = f(x_0)\}$.*

Для того, чтобы запомнить, что чему должно принадлежать, надо вспомнить, что размерность S (и размерность TS_{x_0}) может быть любая от 1 до $n-1$, а размерность поверхности уровня N (и размерность TN_{x_0}) обязательно равна $n-1$.

Доказательство. Возьмем произвольный вектор $\xi \in TS_{x_0}$ и такую гладкую кривую $x(t)$ на поверхности S, что $x'(0) = \xi$. это можно всегда сделать, мы это доказали. Если x_0 — точка локального условного экстремума, то точка 0 — точка локального экстремума гладкой функции $f(x(t))$. Поэтому, необходимо, $(f(x))'_t(0) = 0$, то есть $f'(x_0)x'(0) = (\nabla f(x_0), \xi) = 0$. Так как точка x_0 — не критическая, то есть $\nabla f(x_0) \neq 0$, то это равенство как раз и означает, что $\xi \in TN_{x_0}$. $\square$

Лекция 15 (08 июня 2016)

На прошлой лекции мы каким уравнениям удовлетворяют точки касательного пространства.

Рассмотрели задачу об условном экстремуме и сформулировали геометрическую теорему о необходимом условии.

Напомню коротенько формулы (они нам все равно понадобятся для метода множителей Лагранжа) и теорему.

Формулы мне понадобятся такие. Пусть есть $f : \mathbb{R}^n \to \mathbb{R}$, гладкая. Пусть точка x_0 не критическая, то есть $\nabla f(x_0) \neq 0$. У поверхности уровня $f(x) = f(x_0)$, проходящая через точку x_0 в точке x_0 есть касательная гиперплоскость, её уравнение имеет вид $(\nabla f(x_0), x - x_0) = 0$. Так как в окрестности некритической точки все точки некритические, то то же самое будет и в любой точке из малой окрестности x_0.

Если поверхность S задана $n-k$ уравнениями $F_j(x) = 0$, то совокупность уравнений касательной k мерной плоскости в точке $x_0 \in S$ имеет вид $(\nabla F_j(x_0), x - x_0) = 0$. Или, если считать, что эти уравнения — это уравнение $F(x) = 0$, где $F : \mathbb{R}^n \to \mathbb{R}^{n-k}$, то уравнение касательной плоскости размерности k имеют вид $DF(x_0)(x - x_0) = 0$.

Для того, чтобы точка x_0 была бы точкой условного локального экстремума для функции f, необходимо выполнение условия $TS_{x_0} \subset TN_{x_0}$, где TS_{x_0} — пространство, касательное к S, а TN_{x_0} — пространство, касательное к поверхности уровня $N = \{x \in D : f(x) = f(x_0)\}$.

Прежде чем идти дальше я сформулирую ещё чисто алгебраическое утверждение.

Лемма. *Пусть даны векторы $a_i \in \mathbb{R}^n$ и ещё вектор $b \in \mathbb{R}^n$. Допустим, что $(a_i, x) = 0$ влечёт $(b, x) = 0$ для любого $x \in \mathbb{R}^n$. Тогда b есть линейная комбинация a_i.*

Доказательство. Пусть Π — подпространство, натянутое на a_i. То есть условие теоремы — это: есть $x \perp \Pi$ влечёт $x \perp b$ для любого x, то $b \in \Pi$.

От противного. Пусть $b \notin \Pi$, тогда $b = b_\perp + b_\Pi$, $b_\perp \neq 0$, $b_\perp \perp \Pi$, $b_{Pi} \in \Pi$. Теперь положим $x = b_{bot}$, $(x, a_i) = 0$, но $(x, b) \neq 0$. Противоречие.

Метод множителей Лагранжа.

Теорема с прошлой лекции красиво формулируется и легко доказывается, однако для прямого применения в конкретных задачах она не кажется приемлемой. Зато их неё вытекает замечательный метод практического нахождения точек, подозрительных на условный экстремум — метод множителей Лагранжа.

Пусть поверхность S задана системой из m уравнений

$$\begin{cases} F_1(x_1, \ldots, x_n) & = 0, \\ \quad \ldots\ldots & \\ F_m(x_1, \ldots, x_n) & = 0. \end{cases}$$

Пусть ранг этой системы в окрестности точки x_0 постоянный и равен m, то эта система уравнений задаёт в окрестности точки x_0 гладкую поверхность размерности $k = n - m$.

Тогда пространство TS_{x_0} задаётся системой линейных уравнений

$$\begin{cases} \partial_1 F_1(x_0)\xi_1 & + \ldots + & \partial_n F_1(x_0)\xi_n & = 0, \\ \ldots\ldots\ldots\ldots & & & \\ \partial_1 F_m(x_0)\xi_1 & + \ldots + & \partial_n F_m(x_0)\xi_n & = 0. \end{cases}$$

Пространство TN_{x_0} задается уравнением

$$\partial_1 f(x_0)\xi_1 + \ldots + \partial_n f(x_0)\xi_n = 0.$$

Условия теоремы означают, что это уравнение является следствием системы уравнений. Таким образом, линейная комбинация уравнений системы является вот этим уравнением.

Иначе говоря, при некоторых λ_i справедливо равенство

$$\mathbf{grad}\, f(x_0) = \sum_{i=1}^{m} \lambda_i\, \mathbf{grad}\, F_i(x_0),$$

Это n скалярных равенств $\dfrac{\partial f}{\partial x_k} = \sum\limits_{i=1}^{m} \lambda_i \dfrac{\partial F_i}{\partial x_k},\ k = 1, \ldots, n$. Неизвестных $m+n$, если добавить ещё m равенств, определяющих поверхность, то получится разумно определённая система. Существование таких λ_i, что выполнено такое равенство в $\mathbb{R}^n$, и есть необходимое условие локального условного экстремума.

Обычно делают так. Пишут *функцию Лагранжа*: $L(x, \lambda) = f(x) - \sum\limits_{i=1}^{m} \lambda_i F_i(x)$ от $m+n$ переменных, n переменных x и m переменных λ. Теперь выписываем условия обычного безусловного экстремума для функции L по всем переменным — это как раз будут равенства для градиента и для поверхности:

$$\frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial \lambda_j} = 0, \qquad i = 1, \ldots, n;\ j = 1, \ldots, m.$$

Замечания.

1. Можно метод множителей Лагранжа применять и к случаю $k < n - m$. То есть, не страшно, если будет лишнее условие.

2. Пишут функцию Лагранжа и так: $L(x, \lambda) = f(x) + \sum\limits_{i=1}^{m} \lambda_i F_i(x)$ (с «+» вместо «−»).

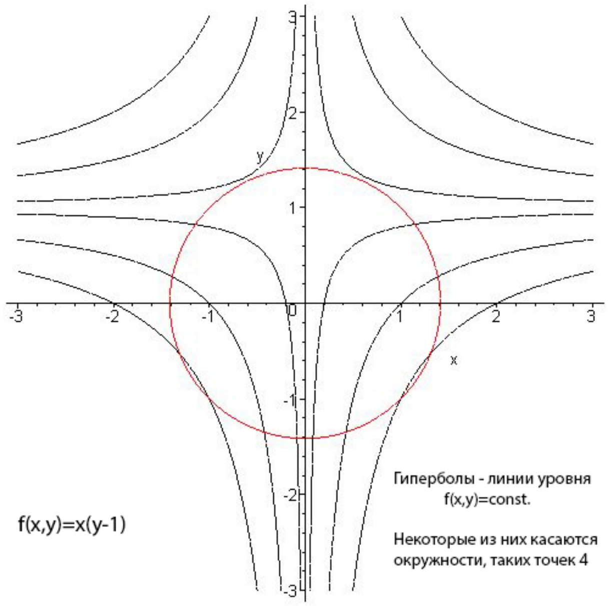


Рис. 7 Пример.
Найти точки условного экстремума функции

$$f(x, y) = x(y - 1)$$

на окружности $x^2 + y^2 = 2$. Пишем функцию Лагранжа

$$L(x, y, \lambda) = x(y - 1) - \lambda(x^2 + y^2 - 2)$$

и ищем её критические точки. Уравнения

$$x^2 + y^2 - 2 = y - 1 + 2\lambda x = x + 2\lambda y = 0.$$

Решаем, получаем 4 решения, 2 условных минимума, 2 условных максимума.

Достаточные условия условного экстремума.

Достаточные условия экстремума даже в случае безусловном были не слишком простые: нужно было проверять знакоопреденность квадратичной формы. Естественно, в случае условного экстремума они не могут быть «проще».

Пусть в $\mathbb{R}^n$ есть точка x_0, пусть в некоторой окрестности $D \ni x_0$ задана функция $f : D \to \mathbb{R} \in C^2$, причем точка x_0 не является критической для f.

Пусть гладкая поверхность S содержит точку x_0. Пусть в этой точке при некоторых λ_j выполнено необходимое условие локального условного экстремума: функция Лагранжа $L(x) = L(x, \lambda)$ имеет критическую точку.

Теорема (достаточные условия условного экстремума). *Для того, чтобы точка x_0 была точкой условного экстремума функции f достаточно, чтобы квадратичная форма $\sum \partial_{ij} L(x_0)\xi_i\xi_j$ была знакоопределенной на TS_{x_0}. Для того, чтобы точка x_0 не была точкой условного экстремума функции f достаточно, чтобы квадратичная форма $\sum \partial_{ij} L(x_0)\xi_i\xi_j$ не была знакоопределенной (нестрого) на TS_{x_0}.*

Как и ранее, случай нестрогой определённости-неопределённости этой формы оставляет вопрос об экстремуме открытым.

Если мы покажем, что функция L имеет условный экстремум в точке x_0, то и функция f имеет там условный экстремум: $f(x) = L(x)$ при $x \in S$.

Повторю: приведенное достаточное условие должно быть выполнено только для векторов из касательного подпространства, вовсе не для любых векторов из большого пространства. В пространстве $\mathbb{R}^n$ (большом) есть маленькое касательное подпространство. Форма определена на n-мерных векторах, принадлежащих k-мерному касательному подпространству.

Доказательство. Так как выполнено необходимое условие экстремума и так как на S частные производные по λ_j равны нулю, то тейлоровское разложение функции $L(x)$ в x_0 имеет вид

$$L(x) = L(x_0) + \frac{1}{2!}\sum_{i,j=1}^{n} \frac{\partial^2 L(x_0)}{\partial x_i \partial x_j}(x_i - x_{i,0})(x_j - x_{j,0}) + o(\|x - x_0\|^2),$$

линейные слагаемые обращаются в ноль на S.

В силу определения поверхности существует гладкое биективное отображение $x(t)$ окрестности нуля в $\mathbb{R}^k$ в $\mathbb{R}^n$ на множество $U\bigcap S$, где $U \in \mathbb{R}^n$ — окрестность точки x_0.

Из соотношения $x(t) = x(0)+x'(0)t+o(\|t\|)$ (здесь $t \in \mathbb{R}^k$, $x'(0) \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^n)$, $x \in \mathbb{R}^n$, $x(0) = x_0$) следует соотношение

$$\|x(t) - x(0)\|_{\mathbb{R}^n} \sim \|t\|_{\mathbb{R}^k}.$$

Так как $x_i - x_{i,0} = \sum\limits_{r=1}^{k} \partial_r x_i(0) t_r + o(\|t\|)$, и $x_j - x_{j,0} = \sum\limits_{s=1}^{k} \partial_s x_i(0) t_s + o(\|t\|)$, то

$$(x_i - x_{i,0})(x_j - x_{j,0}) = \sum_{r,s=1}^{k} \partial_s x_i(0) \partial_r x_j(0) t_r t_s + o(\|t\|^2).$$

Таким образом,

$$L(x(t)) - L(x(0)) = \frac{1}{2!}\sum_{i,j=1}^{n} \frac{\partial^2 L(x_0)}{\partial x_i \partial x_j} \sum_{r,s=1}^{k} \partial_s x_i(0) \partial_r x_j(0) t_r t_s + o(\|t\|^2).$$

Осталось заметить, что квадратичная форма в правой части — это другая форма записи формы

$$\sum_{i,j=1}^{n} \frac{\partial^2 L(x_0)}{\partial x_i(0) \partial x_j(0)} \xi_i \xi_j$$

при $\xi \in S$. Тогда все будет доказано: знак выражения $L(x(t)) - L(x(0))$ совпадает со знаком квадратичной формы.

Совпадение форм следует из равенств $\xi_i = \sum_{s=1}^{k} \partial_s x_i(0) t_s$. Эти равенства следуют из параметрической записи $\xi = x'(0)t$ для векторов касательного пространства TS_{x_0}. $\square$

Замечания.

1. Задача на условный экстремум: $f(x, y) = x^2 + y^2,\ \ x^2/4 + y^2/9 \leqslant 1$.

Это простая задача: сначала надо поискать экстремумы во внутренних точках эллипса. Для этого нужно приравнять к нулю градиент f, естественно, мы найдем минимум в точке $(0, 0)$.

Теперь нужно пытаться найти «интересные» точки на границе эллипса.

Таких точек будет 2: $(0, 3)$ и $(2, 0)$. В этих точках линии уровня f — окружности с центром в нуле касаются границы эллипса. При этом в точке $(2, 0)$ касание внутреннее, там точка, подозрительная на минимум, и это в самом деле будет минимум на границе, но внутри эллипса будут меньшие значения. А в точке $(0,3)$ касание внешнее, эта точка подозрительна на максимум, и является максимумом.

Ответ: есть точка $(0,0)$ — минимум, точка $(0,3)$ — максимум. Одна точка внутри области, другая на границе.

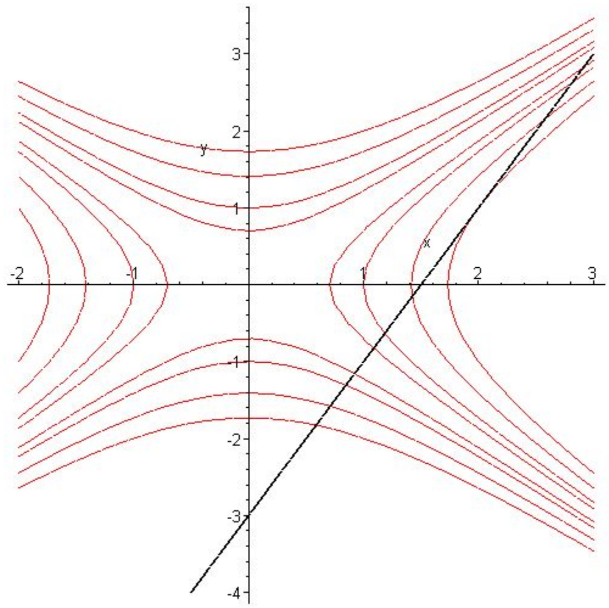


Рис. 8 Игрушечный пример.
$f(x,y)=x^2-y^2$, $2x-y-3=0$.
Функция Лагранжа
$L(x,\lambda)=x^2-y^2-\lambda(2x-y-3)$.
Уравнения подозрительных точек:
$2x-y-3=2x-2\lambda=-2y+\lambda=0$.
Решение: $x=2,\ y=1,\ \lambda=2$.
Форма имеет вид: $\xi_1^2-\xi_2^2$, уравнение касательной: $2\xi_1=\xi_2$, поэтому на касательном пространстве форма имеет вид $-3(\xi_2)^2$, то есть это точка максимума. Такой пример легко решить и без Лагранжа. Обратите внимание, что форма на всем пространстве не является знакоопределённой, а на касательной становится знакоопределённой.

2. Задача об условном экстремуме рассматривается не только на гладких многообразиях, да и не только гладкие функции возникают в практических приложениях. Сказать про негладкие теоремы. Что бывает выпуклый анализ, негладкий анализ.

Рассказать про принцип максимума, про вариационное исчисление, про уравнения Эйлера–Лагранжа, про бесконечномерные задачи.

Задача о брахистохроне — кривой скорейшего спуска. Задача о её нахождении была поставлена в 1696 году Иоганном Бернулли.

Среди плоских кривых, соединяющих две данные точки A и B, лежащих в одной вертикальной плоскости (B ниже A), найти ту, двигаясь по которой под действием только силы тяжести, сонаправленной отрицательной полуоси OY, материальная точка достигнет B из A за кратчайшее время.

Метод решения, полученного 26 января 1697 года Ньютоном, лег в основу важнейшей области естествознания — вариационного исчисления.

Решением задачи о брахистохроне является дуга циклоиды с горизонтальным основанием, точка возврата которой находится в точке A, или иными словами, имеющая вертикальную касательную в точке A.

3. Сказать что-то про задачу линейного программирования.